

Software Selegen-REML/BLUP: a useful tool for plant breeding

Marcos Deon Vilela de Resende^{1*}

Crop Breeding and Applied Biotechnology
16: 330 -339, 2016
Brazilian Society of Plant Breeding.
Printed in Brazil
<http://dx.doi.org/10.1590/1984-70332016v16n4a49>

Abstract: *The software Selegen-REML/BLUP uses mixed models, and was developed to optimize the routine of plant breeding programs. It addresses the following plants categories: allogamous, automagous, of mixed mating system, and of clonal propagation. It considers several experimental designs, mating designs, genotype x environment interaction, experiments repeated over sites, repeated measures, progenies belonging to several populations, among other factors. The software adjusts effects, estimates variance components, genetic additive, dominance and genotypic values of individuals, genetic gain with selection, effective population size, and other parameters of interest to plant breeding. It allows testing the significance of the effects by means of likelihood ratio test (LRT) and analysis of deviance. It addresses continuous variables (linear models) and categorical variables (generalized linear models). Selegen-REML/BLUP is friendly, easy to use and interpret, and allows dealing efficiently with most of the situations in plant breeding. It is free and available at http://www.det.ufv.br/ppestbio/corpo_docente.php under the author's name.*

Key words: *Linear mixed models, prediction, variance components, BLUP, REML, selection.*

INTRODUCTION

The estimation of variance components and the prediction of genetic values are essential procedures in breeding programs. In the 1990s, there was a qualitative progress in the analytical methodologies of genetic parameters estimation and selection applied to plant breeding. Currently, REML/BLUP (Residual or Restricted Maximum Likelihood/Best Linear Unbiased Prediction) is the standard procedure for the estimation of genetic parameters and optimal selection in several species.

The field trial, as a rule, is associated with imbalance of data due to several reasons, such as loss of plants and plots, unequal quantities of seeds and seedlings available by treatment, experimental net with different numbers of replications per experiment, and different experimental designs, non-evaluation of all combinations of genotypes and environments, among others. As a result, REML/BLUP method is an optimal procedure of genotypic evaluation, and it is also known as mixed model methodology. This procedure naturally deals with the imbalance, leading to more accurate estimates and predictions of genetic parameters and genetic values, respectively.

BLUP is the optimal selection procedure for additive genetic effects (a), of dominance effects (d) and genotypic effects (g), depending on the situation.

***Corresponding author:**
E-mail: marcos.deon@gmail.com

Received: 01 September 2016
Accepted: 30 September 2016

¹ Embrapa Florestas, Universidade Federal de Viçosa, Departamento de Engenharia Florestal, 36.570-000, Viçosa, MG, Brazil

BLUP maximizes selective accuracy, and allows the simultaneous use of several information sources, such as those from various experiments installed in one or several locations and evaluated in one or more crops. The individual BLUP uses all the effects of the statistical model, addresses imbalance, considers the genetic relatedness between the evaluated plants, and the coincidence between selection and recombination units.

The main practical advantages of using REML/BLUP is that they: allow comparing individuals or varieties over time (generations, years) and space (location, blocks); allow simultaneous correction for environmental effects, estimation of variance components, and prediction of genetic values; allows dealing with complex data structures (repeated measurements, different years, location and designs); may be applied to unbalanced data and to not orthogonal designs.

At the end of the second millennium, pioneering works in Brazil applied to plant breeding were carried out in the field of mixed linear models fitted under the frequentist approach via REML/BLUP (Resende et al. 1993, 1996, Bueno Filho and Vencovsky 2000, Duarte and Vencovsky 2001), and the Bayesian approach via Gibbs sampler (Resende 1997, Resende and Rosa-Perez 1999, Resende 2000). These models included random regressions for longitudinal and multivariate data (Resende 1997, Resende and Rosa-Perez 1999), non-linear or generalized linear models for categorical variables (Resende 2000), spatial analysis (Resende and Sturion 2001), factor analytical mixed models for multivariate analysis and genotype x environment interaction (Resende and Thompson 2003, 2004), and competition or associative models of social interaction (Resende and Thompson 2003, Resende et al. 2005). Pioneering works with Genomic Selection were also carried out (Resende 2007, Resende et al. 2008). These works are summarized in Resende (2002, 2007, 2015) and Resende et al. (2014), which also include reaction norms, structural equations and survival analysis models for censored data.

The generic theory of BLUP as optimal procedure was widespread from the 1970s by the scientists Charles Henderson, in the United States (Henderson 1973, 1975, 1976) and Robin Thompson, in England (Thompson 1976, 1977, 1979), among others. For the application of BLUP, it is necessary reliable estimates of variance components. REML is the optimal method of estimation of variance components, with unbalanced or not data, which was developed by Patterson and Thompson (1971) and Thompson (1973, 1977, 1980).

The REML/BLUP procedure became popular abroad in animal breeding from the 1980s. In Brazil, the method began to be used in dairy cattle from 1994 (Verneque and Valente 2001). This was due to the development of specific softwares that allow the proper handling of additive genetic relationship matrix among the evaluated individuals. Algorithms to directly write this additive genetic relationship matrix were presented by Henderson (1976), in the USA and by Thompson (1977), in England. The software Selegen-REML/BLUP (Statistical system and computerized genetic selection via linear mixed models) was created in association with the improvement of the genetic selection methodologies from the mathematical and statistical analysis of field experiments data. Prior to the emergence of the softwares ASREML and SAS, for mixed models, which were created 1996, Selegen was released in 1993, together with the use of the mixed linear models (via best unbiased linear prediction) in plant breeding in Brazil, which is currently widely used (Bueno Filho and Vencovsky 2009).

The software Selegen-REML/BLUP and its applications

The software Selegen-REML/BLUP was developed to meet the demands of plant breeding programs, and includes the following categories of plants: allogamous, autogamous, of mixed mating system, and of clonal propagation. It considers various experimental designs, several mating designs, genotype x environment interaction, experiments repeated over sites, repeated measures, progenies belonging to several populations, among other factors. The software not only fits the effects and presents the variance components, but also shows the additive genetic values, the dominance genetic values, and the genotypic genetic values of the individuals, the genetic gain with selection, the effective population size, among other parameters of interest in plant breeding. From a statistical point of view, it is also interesting, since it allows testing the significance of the effects by means of the likelihood ratio test (LRT) and analysis of deviance. It also addresses continuous variables (linear models) and categorical variables (generalized linear models).

Selegen-REML / BLUP is easy to use and interpret, and allows dealing efficiently with the most common situations in plant breeding. It is of free access in universities and public research institutes in Brazil and abroad. In the private sector, it has been used in the breeding of coffee, forage crops, fruit trees, forest trees, eucalyptus, pine, black wattle, teak, corn, soybeans, rice,

rubber tree, sugarcane, among other species. In public institutions, it has been used in these same species and also in beans, cashew, acerola, cupuaçu, cocoa, coffee, guarana, palm, peach palm, royal palm, orange, brachiaria, panicum, stylosanthes, leucaena, mate tea, pequi, potato, cassava, açaí, mango, passion fruit, camu-camu, buriti, among others.

Mathematical and computational algorithm

The computational implementation of the mixed model methodology is based on heavily numerical methods, especially in numerical linear algebra in order to obtain iterative solutions of mixed model equations (obtainment of BLUP) and numerical calculus for the maximization/minimization of the functions of several variables, in order to obtain the REML estimates.

Several computational algorithms for the obtainment of variance components by REML have been developed, such as MS (Fisher's Method of Scoring), EM (Expectation-Maximization), and AI-REML (Average Information-REML). The EM algorithm is numerically very stable, showing convergence even if initial values are not entirely adequate. However, its disadvantage is the slowness for estimates close to the limit of the parametric space (for instance, when a variance tends to zero). If positive initial values are used, the convergence to non-negative values is guaranteed (Harville 1977).

The EM algorithm works by means of the obtainment of the expectation (by integration) and maximization (derivation) of the likelihood function of the data, successively. In models of individual plants, in which the order of the mixed model equations usually exceeds the number of observations, the obtainment of estimates by means of first derivative by the EM method requires the inversion of the coefficient matrix of the mixed model equations, increasing computational effort. The methods of Newton-Raphson and of Fisher have quadratic convergence, whereas the EM algorithm presents linear convergence, and is therefore slower.

The AI-REML algorithm is three times faster than the EM, and is an improved derivative procedure that uses the first and second derivative of the likelihood function. This algorithm is based on the use of the information of the mean of the observed and expected second derivatives of the likelihood function, so that the term containing traces of the inverse matrix product is canceled, leaving a simpler expression for computation. Sparse matrices techniques are used in the calculation of the elements of the inverse matrix of the coefficients, which are necessary for the first derivatives of the likelihood function.

However, sometimes the AI algorithm failures in the convergence, and therefore, the EM algorithm would be an alternative, which ensures an increase in the log L in each iteration. The use of the proposed PX-EM algorithm (Parameter expanded EM) is an excellent choice, and it presents stability and good performance. The software Selegen-REML/BLUP combines the methods of Takahashi and the method of Zollenkopf sparse bifactorization in the EM algorithm with PX-EM. Therefore, the algorithm is PX-EM-SB-type (Expanded Parameter Expectation-Maximization/Sparse Bifactorization), and the software is the PX-EM-SB-REML-type.

The methods of Takahashi and Zollenkopf were developed in Electrical Engineering, associated with matrices of impedance in electrical circuits. Both methods are similar, and the Takahashi algorithm is naturally generated by multiplication factors to the left and to the right of the Zollenkopf bifactorization in reverse order. Estimators and predictors used in Selegen-REML/BLUP are described in a series of 10 papers published by Resende in the Journal of Mathematics and Statistics, from 1999 to 2006, and in Resende et al. (1996, 2014) and Resende (1999, 2000, 2002).

The restricted likelihood function to be maximized is given below. A generalized linear mixed model is given by $y = X\beta + Z\tau + \varepsilon$, with the following distributions and structures of means and variances:

$$\begin{aligned} \tau &\sim N(0, G) & E(y) &= X\beta \\ \varepsilon &\sim N(0, R) & \text{Var}(y) &= V = ZGZ' + R \end{aligned}$$

In which:

y : known vector of observations

β : parametric vector of fixed effects, with incidence matrix X.

τ : parametric vector of random effects, with incidence matrix Z .

ε : unknown vector of errors.

G : variance-covariance matrix of random effects.

R : variance-covariance matrix of errors.

0 : null vector.

Assuming G and R as known, the simultaneous estimation of fixed effects and the prediction of random effects can be obtained by means of mixed model equations (BLUP) given by

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\tau} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

When G and R are not known, the variance components associated with random effects can be efficiently estimated by the REML method (Patterson and Thompson 1971). Except for a constant, the residual likelihood function (in terms of its log) to be maximized is given by:

$$\begin{aligned} L &= -\frac{1}{2} (\log|X'V^{-1}X| + \log|V| + \nu \log \sigma_{\varepsilon}^2 + y'Py/\sigma_{\varepsilon}^2) \\ &= -\frac{1}{2} (\log|C^*| + \log|R| + \log|G| + \nu \log \sigma_{\varepsilon}^2 + y'Py/\sigma_{\varepsilon}^2) \end{aligned}$$

in which:

$$V = R + ZGZ'; \quad P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}.$$

$\nu = N - r(x)$: degrees of freedom for the random effects, in which N is the total number of data, and $r(x)$ is the rank of the matrix X .

C^* = matrix of the coefficients of the mixed model equations.

Overall, the generalized model described includes several unique models to each situation.

Statistical, mathematical and genetic procedures of Selegen-REML/BLUP

Selegen-REML/BLUP was developed in Fortran 90 language, and has Windows and DOS operating systems interface. It is suitable for analysis of either balanced or unbalanced experiments, leading to maximum efficiency. There is no need to inform whether the experiment is balanced or not, since it uses optimal and generic mathematical and statistical procedure for any situation.

The files to be analyzed must have .txt extension (MS-DOS text), with a header line. This line is only for user guidance, it is ignored by the program. Thus, the line may contain any names. The program uses classificatory and dependent variables with up to 15 digits. However, for reasons of space, in the program's outputs, the classificatory variables are presented with up to seven digits only. It is recommended the use up to seven decimal digits for these variables. Such classificatory variables may be alpha-numeric.

In statistical terms, the following methods of analysis are performed by Selegen-REML/BLUP:

- General statistics: mean, variance, standard deviation, coefficient of variation, maximum, minimum, kurtosis, asymmetry, covariance, correlation and commonality
- Analysis of variance
- Analysis of covariance and correlation
- Multivariate analysis: principal components and cluster analysis
- Mixed linear models via REML/BLUP

MDV Resende

- Mixed linear models via REML/BLUE
- Generalized linear mixed models via REML/BLUP
- Linear mixed models for repeated measures
- Linear mixed models for multiple experiments analysis and G x E interaction
- Mixed linear models with covariates
- Maximization of residual likelihood function for likelihood ratio test (LRT) and analysis of deviance
- Mixed linear models with residual variance heterogeneity within treatments
- Point and interval prediction of random effects (breeding values)
- Computation of residuals for analysis of homogeneity and normality
- Spatial autocorrelation for analysis of residuals
- Hierarchical analysis for population genetics
- Genetic sampling (effective population size)

The current program includes about 200 models of analysis, and in terms of experimentation and genetic improvement, it provides the following results of interest:

- BLUP and REML/BLUP for additive, dominance and genotypic effects
- Heritabilities, genetic and phenotypic correlations, genetic gain
- BLUP under heterogeneity of variance within treatments
- Analysis of deviance
- Experimental designs: randomized, complete blocks, augmented blocks, lattice, row and column
- Groups of experiments, several locations, and genotype x environment interaction
- Mating designs: open and controlled pollination (half-sib, full-sib, factorial, diallel, hierarchical, unbalanced designs, hybrids and selfing generations)
- Clonal tests
- One or more populations
- One or more plants per plot
- One or more repeated measures
- Principal genetic components
- Cluster analysis by genotypic values
- Genetic divergence via genotypic values
- Multi-trait selection index
- Effective population size
- Optimization of selection and restriction in inbreeding
- Stability and adaptability of genotypic values
- Population genetics
- Allogamous and autogamous species, and of mixed mating system, animals

- Associated selection methods: selection of parents (ag), selection of potentials parents (a), selection of potential clones ($g = a + d$), selection of crosses ($v_{gc} = 0.5 (af + am) + cec$), selection of clones. The Selegen-Reml/Blup was designed to maximize the overall efficiency of breeding, and addresses in an intricate way the topics mentioned above, overlapping recurrent selection scheme, crossing design, experimental design, statistical control via covariate, and propagation system of improved material.

Genetic evaluation models in function of the species and of the experimental structure

In plant breeding, there are certain types of cultivars (Table 1). The three types of cultivars considered in this study aim at capturing the genotypes with the two most important characteristics for the production system (sustainable yield and products homogeneity), given the biological conditions (reproduction and propagation systems) of each species. Thus, the breeding strategies (which generate the data to be statistically analyzed) are similar for the species within each of these three types. Clones and hybrids between inbred lines capture both additive genetic effects and dominance effects (which, in presence of genetic diversity, provide heterosis). Inbred lines capture only the additive genetic effects. Breeding programs aiming at inbred lines are similar (hybridization between lines, followed by inbred generations conduction until the selection of new lines) to those aiming at hybrids. However, when the hybrids are the target, the process continues with the cross between lines, in order to identify the superior hybrids. With the advent of double-haploid, the stage of inbred generations conduction has been suppressed, and the obtained lines (so far not tested in the field), have their crosses predicted with the aid of the phenomics and genomics.

Table 1. Types of cultivars considered in plant breeding

Types of cultivars	Mating system of species	Examples
Clones	Allogamous, Autogamous, Asexuals	Perennial and allogamous (sugarcane, eucalyptus, guarana, rubber tree, cocoa, canephora coffee, fruit trees, forage, etc.); Perennial and autogamous (peach); Perennial with annual cycle (cassava, potato)
Lines	Autogamous	Annual and Autogamous (rice, wheat, oat, beans, soybeans, lettuce, pepper); Perennial and autogamous (Coffee arabica)
Hybrids of Lines	Allogamous, Autogamous	Annual and allogamous annual (corn, sunflower, onion, carrot, brassicas, pumpkin); Annual and autogamous (tomato, pepper, tobacco)

General models of genetic evaluation, which can be applied to plant breeding are presented below. The procedures are similar within types, according to the following classification.

a) **Perennial and allogamous plants** (eucalyptus, pine, canephora coffee, sugarcane, fruit trees, forage): Estimation is simultaneously based on heritability, repeatability, covariance structure, and growth curves; genealogy is variable.

b) **Perennial and autogamous plants** (Coffee arabica, peach, apricot, dwarf coconut, lemon, leucaena): Estimation is simultaneously based on heritability, repeatability, covariance structure, and growth curves; genealogy is fixed from each F2 generation to F6 generation.

c) **Perennial plants with asexual reproduction** (sugarcane, rubber tree, orange, brachiaria, panicum): Estimation is simultaneously based on broad-sense heritability, repeatability, covariance structure, and growth curves; genealogy is constant through clonal stages

d) **Annual and Allogamous** (corn, popcorn, sunflower, broccoli, carrots): Estimation is based on heritability; genealogy is variable.

e) **Annual and autogamous plants** (beans, soybeans, rice, wheat, oat): Estimation is based on the heritability; genealogy is fixed from each F2 generation to F6 generation.

f) **Annual plants with asexual reproduction** (cassava, potato): Estimates based on broad-sense heritability; genealogy is constant through clonal stages

A rough classification of groups of plants regarding the models of genetic evaluation in function of the experimental structure in the field and models of genetic evaluation is presented below:

a) **Annual and vegetable crops** (rice, corn, soybeans, wheat, oat, barley, sorghum, cotton, potato, cassava): data analyzed with one observation per plot and inference at the level of genetic treatment effects (lines, hybrids, clones, families, cultivars, accessions)

b) **Forage and sugarcane**: data analyzed with one observation per plot and repeated measures, inference at level of genetic treatments effects (lines, hybrids, clones, families, cultivars, accessions)

c) **Forest species**: data taken at the level of individual plants, without repeated measures, inference to the level of individuals, parents and clones

d) **Fruit trees, palm trees** (açai, coconut, palm, peach palm, date palm), stimulants (coffee, cocoa, guarana, mate tea), rubber tree: data collected from individual plants, with repeated measures, inference at the level of individuals, parents and clones

The complexity of the models, the difficulty in the genetic evaluation, and the imbalance degree increase from (a) to (d), i.e., complexity increases in the following order: annual and vegetable crops; forage and sugarcane; forest species; fruit trees.

Significance of the effects of the model and complete statistical analysis

Genotypic evaluations addresses the estimation of variance components (genetic parameters) and the prediction of genotypic values. Estimates of genetic parameters, such as heritability and genetic correlations, are fundamental to the design of efficient breeding strategies. REML is an efficient method in the study of the several variation sources associated with the evaluation of field experiments, and allows decomposing the phenotypic variation in their several genetic and environmental components and in genotype x environment interaction components.

In the mixed models analysis with unbalanced data, the effects of the model are not tested via F tests, as it is done in the method of analysis of variance. In this case, for the random effects, likelihood ratio test is recommended (LRT). For fixed effects, an approximate F test can be used. A similar Table to that of analysis of variance can be elaborated. Such table can be called Analysis of deviance (ANADEV), and is established according to the following steps:

a) obtainment of the maximum point of the logarithm of the residual likelihood function (L) for models with and without the effect to be tested

b) obtainment of the deviance $D = -2\text{Log}L$ for models with and without the effect to be tested

c) differentiation between deviances for models with and without the effect to be tested, with the obtainment of the likelihood ratio (LR)

d) testing, via LRT, of the significance of this difference using the chi-square test with 1 or 0.5 degrees of freedom.

As an example, the following experiment was carried out in randomized blocks with several plants per plot, and the following model was specified: $y = \mu + g + b + gb + e$, in which g is the vector of random effects of genotypes, b is the vector of fixed effects of blocks, gb is the vector of random effects of plots, and e is the vector of random residuals within plots. The following analysis of deviance (ANADEV) can be carried out (Table 2).

The software Selegen provides (.dev extension files) the deviances when it is processed models with or without (just by resetting to zero the corresponding coefficient of determination c2 on Selegen's display) the effects to be tested. With

Table 2. Analysis of deviance (ANADEV)

Effect	Deviance	LRT(chi-square)	Variance components	Coefficient of determination
Genotypes	647.18	6.55**	0.0329*	$h^2_{\text{genotype}} = 0.045^*$
Plots	654.13	13.50**	0.0685**	$c^2_{\text{plot}} = 0.094^{**}$
Residuals	-	-	0.6206	$c^2_{\text{residual}} = 0.859$
Compleat model	640.62	-	-	$c^2_{\text{total}} = 1.000$
Block	-	$F = 7.012^{**}$	-	-

Chi-square of 3.84 and 6.63 for 5% and 1% significance, respectively

these deviances, it becomes easy to build the table of analysis of deviance. In the present example, it is verified that the effects of genotype and plots are significant. Consequently, the respective variance components are significantly different from zero, as well as their coefficients of determination (heritability of genotypic effects - h^2_{genotype} and coefficient of determination of the effects of plots - c^2_{plot} , as obtained by the models 1 and 2 of Selegen). The factor block, considered as fixed effect, was tested by the F test of Snedecor.

In general, a complete statistical analysis involves the following six activities: the estimation of mean components; the estimation of variance components; hypothesis tests; the inference regarding the accuracy (reliability); the bias; and precision of the estimation/prediction. Considering the mixed models, these activities involve BLUP prediction, REML estimation, analysis of deviance, calculation of the prediction accuracy and of the variance of the prediction error, respectively. In the REML/BLUP procedure, bias is assumed as zero, since these estimators/predictors belong to the class of the best linear unbiased estimators/predictors (BLUE/BLUP).

Results generated by the six activities of a complete statistical analysis should be discussed in the papers. In the genetic field, the following script should be followed:

Hypothesis test: inferences on the significance of the genetic variability (V_g), using the analysis of deviance or the F test of the analysis of variance

Variance components and their proportions: inferences on the genetic control (high, moderate, and low, according to Resende 2002), or heritabilities and correlations between traits, coefficients of variation, repeatabilities

Mean components: genetic values and genetic gain

Precision: PEV (variance of the prediction error); ratio PEV/ V_g (with parametric space between 0 and 1)

Accuracy: with parametric space between 0 and 1, and classification according to Resende and Duarte (2007)

Bias: given by the regression of y in \hat{y} ($\beta(y/\hat{y})$), in which $\beta = 1$ is the ideal and indicates unbiasedness.

Moreover, studies on diversity (population effective size or N_e ; genetic distances, and multivariate groupings) complement the inferences.

Interaction between organisms

There are also estimates based on genotypes of two organisms in a single individual, such as those of rubber tree, orange, peach, mango, which are cultivated in rootstock + graft. This also occurs in the experiments involving plant x pathogen interaction. This type of statistical analysis is addressed in the software Selegen-REML/BLUP. The study on plant x pathogen interactions involves the evaluation of different genotypes (accessions) of the plant species subjected to inoculation with different races, strains or inoculum sources of the microorganisms species (bacteria, fungi, nematodes). Each plant must be inoculated with only one type of microorganism species. The experiment must contain replications of each plant-inoculum combination. Thus, the analysis model will contain the effects of replications (b, for blocks designs), genotypic effects (g) of the plant accessions, genotypic effects (m) of the microorganism strains, plant x pathogen (gm) genotypic interaction, and residual (e). This model, for a variable in a vector y , is given by $y = Xb + Zg + Wm + Tgm + \varepsilon$, in which X, Z, W and T are incidence matrices for the respective effects vectors. Thus, the genetic cause of the disease can be attributed to three factors:

a) Effect on the phenotype, explained by the pathogen genotype

b) Effect on the phenotype, explained by the plant genotype

c) Effect on the phenotype, explained simultaneously by the pathogen and the plant, i.e., the combination of pathogen genotype – plant genotype (is the actual plant – pathogen interaction, or specific combining ability – SCA). The breeding values predicted by the model, for each combination of pathogen genotype – plant genotype, are given by $VGC = \hat{u} + \hat{g}_i + \hat{m}_j + \hat{g}m_{ij}$.

The software Selegen-REML/BLUP has been used internationally (Colombari et al. 2013, Oliveira et al. 2012, Pedrozo et al. 2011, Rosado et al. 2010, Dunlop et al. 2005), and is a very useful tool for plant breeding.

REFERENCES

- Bueno Filho JSS and Vencovsky R (2000) Alternativas de análise de ensaios em látice no melhoramento vegetal. **Pesquisa Agropecuária Brasileira** **35**: 259-296.
- Bueno Filho JSS and Vencovsky R (2009) Selection in several environments by BLP as an alternative to pooled anova in crop breeding. **Ciência e Agrotecnologia** **33**: 1342-1350.
- Colombari Filho JM, Resende MDV, Morais OP, Castro AP, Guimarães EP, Pereira JA, Utumi MM and Breseghello F (2013) Upland rice breeding in Brazil: a simultaneous genotypic evaluation of stability, adaptability and grain yield. **Euphytica** **192**: 117-129.
- Duarte JB and Vencovsky R (2001) Estimação e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. **Scientia Agrícola** **58**: 109-117.
- Dunlop RW, Resende MDV and Beck SL (2005) Early assessment of first year height data from five *Acacia mearnsii* (black wattle) sub-populations in South Africa using REML/BLUP. **Silvae Genetica** **54**: 166-174.
- Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistical Association** **72**: 320-328.
- Henderson CR (1973) Sire evaluation and genetic trends. In **Animal breeding and genetics symposium in honour of J. Lush**. American Society of Animal Science, Champaign, p. 10-41.
- Henderson CR (1975) Use of all relatives in intraherd prediction of breeding values and producing abilities. **Journal of Dairy Science** **58**: 1910-1916.
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. **Biometrics** **31**: 423-449.
- Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. **Biometrics** **32**: 69-83.
- Oliveira EJ, Resende MDV, Santos VS, Ferreira CF, Oliveira GAF, Silva MS, Oliveira, LA and Aguilar-Vildoso CI (2012) Genome-wide selection in cassava. **Euphytica** **187**: 263-276.
- Patterson HD and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. **Biometrika** **58**: 545-554.
- Pedrozo CA, Barbosa MHP, Silva FL, Resende MDV and Peternelli LA (2011) Repeatability of full-sib sugarcane families across harvests and the efficiency of early selection. **Euphytica** **182**: 423-430.
- Resende MDV (1997) Avanços da genética biométrica florestal. In Bandel G, Vello NA and Miranda Filho JB (eds) **Encontro sobre temas de genética e melhoramento: genética biométrica vegetal**. Esalq/Usq, Piracicaba, p. 20-46.
- Resende MDV (2000) **Inferência bayesiana e simulação estocástica (amostragem de Gibbs) na estimação de componentes de variância e valores genéticos em plantas perenes**. Embrapa Florestas, Colombo, 68p.
- Resende MDV (2000) **Análise estatística de modelos mistos via REML/BLUP no melhoramento de plantas perenes**. Embrapa Florestas, Colombo, 101p.
- Resende MDV (2002) **Genética biométrica e estatística no melhoramento de plantas perenes**. Embrapa Informação Tecnológica, Brasília, 975p.
- Resende MDV (2007) Seleção genômica ampla (GWS) e modelos lineares mistos. In Resende MDV (ed) **Matemática e estatística na análise de experimentos e no melhoramento genético**. Embrapa Florestas, Colombo, p. 517-534.
- Resende MDV (2015) **Genética quantitativa e de populações**. Suprema, Visconde do Rio Branco, 452p.
- Resende MDV and Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical** **37**: 182-194.
- Resende MDV, Higa AR and Lavoranti OJ (1993) Predição de valores genéticos no melhoramento de Eucalyptus – melhor predição linear (BLP). **Silvicultura** **43**: 144-147.
- Resende MDV, Lopes PS, Silva RL and Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira** **56**: 63-78.
- Resende MDV, Prates DF, Jesus A and Yamada CK (1996) Estimação de componentes de variância e predição de valores genéticos pelo método da máxima verossimilhança restrita (REML) e melhor predição linear não viciada (BLUP) em Pinus. **Pesquisa Florestal Brasileira** **32/33**: 18-45.
- Resende MDV and Rosa-Perez JRH (1999) **Genética quantitativa e estatística no melhoramento animal**. Editora UFPR, Curitiba, 494p.
- Resende MDV, Silva FFE and Azevedo CF (2014) **Estatística matemática, biométrica e computacional**. Suprema, Visconde do Rio Branco, 881p.
- Resende MDV and Sturion JA (2001) **Análise genética de dados com dependência espacial e temporal no melhoramento de plantas perenes via modelos geoestatísticos e de series temporais empregando REML/BLUP ao nível individual**. Embrapa Florestas, Colombo, 80p.
- Resende MDV, Stringer JK, Cullis BC and Thompson R (2005) Joint modelling of competition and spatial variability in forest field trials. **Brazilian Journal of Mathematics and Statistics** **23**: 7-22.
- Resende MDV and Thompson R (2003) **Multivariate spatial statistical analysis of multiple experiments and longitudinal data**. Embrapa Florestas, Colombo, 126p. (Documento 90).
- Resende MDV and Thompson R (2004) Factor analytic multiplicative mixed models in the analysis of multiple experiments. **Revista de Matemática e Estatística** **22**: 1-22.
- Rosado CCG, Guimarães LMS, Titon M, Lau D, Rosse LN, Resende MDV and Alfenas AC (2010) Resistance to ceratocystis wilt (*Ceratocystis*

Software Selegen-REML/BLUP: a useful tool for plant breeding

- fimbriata*) in parents and progenies of *Eucalyptus grandis* x *E. urophylla*. **Silvae Genetica** **59**: 99-106.
- Thompson R (1973) The estimation of variance and covariance components when records are subject to culling. **Biometrics** **29**: 527-550.
- Thompson R (1976) Relationship between the cumulative difference and best linear unbiased predictor methods of evaluating bulls. **Animal Production** **23**: 15-24.
- Thompson R (1977) The estimation of heritability with unbalanced data. **Biometrics** **33**: 485-504.
- Thompson R (1979) Sire evaluation. **Biometrics** **35**: 339-353.
- Thompson R (1980) Maximum likelihood estimation of variance components. **Mathematische Operationsforschung und Statistik. Series Statistics** **11**: 545-561.
- Verneque RS and Valente J (2001) Avaliação genética de vacas e touros. In Valente J, Durães MC, Martinez ML and Teixeira NM (eds) **Melhoramento genético de bovinos de leite**. Vol. 1, Embrapa Gado de Corte, Juiz de Fora, p. 127-154.