

SELECTION IN SEVERAL ENVIRONMENTS BY BLP AS AN ALTERNATIVE TO POOLED ANOVA IN CROP BREEDING

Seleção em diversos ambientes pelo BLP como alternativa à anava conjunta

Júlio Sílvio de Sousa Bueno Filho¹, Roland Vencovsky²

ABSTRACT

Plant breeders often carry out genetic trials in balanced designs. That is not always the case with animal genetic trials. In plant breeding is usual to select progenies tested in several environments by pooled analysis of variance (ANOVA). This procedure is based on the global averages for each family, although genetic values of progenies are better viewed as random effects. Thus, the appropriate form of analysis is more likely to follow the mixed models approach to progeny tests, which became a common practice in animal breeding. Best Linear Unbiased Prediction (BLUP) is not a "method" but a feature of mixed model estimators (predictors) of random effects and may be derived in so many ways that it has the potential of unifying the statistical theory of linear models (ROBINSON, 1991). When estimates of fixed effects are present is possible to combine information from several different tests by simplifying BLUP, in these situations BLP also has unbiased properties and this lead to BLUP from straightforward heuristics. In this paper some advantages of BLP applied to plant breeding are discussed. Our focus is on how to deal with estimates of progeny means and variances from many environments to work out predictions that have "best" properties (minimum variance linear combinations of progenies' averages). A practical rule for relative weighting is worked out.

Index terms: BLP, plant breeding, statistical genetics.

RESUMO

Os melhoristas de plantas em geral conduzem testes genéticos em delineamentos balanceados, ao contrário do que ocorre com o melhoramento animal. É possível selecionar progênie pela ANAVA conjunta, com base nas médias gerais de cada família. Sabe-se, no entanto, que os valores genéticos de progênie são melhor representados por efeitos aleatórios. As formas de análise dos testes de progênie que parecem mais apropriadas são as que seguem a metodologia de modelos mistos, como no melhoramento animal. Segundo Robinson (1991) o Melhor Preditor Linear Não-Viesado (do inglês, BLUP) não é um método, mas uma propriedade dos estimadores (preditores) dos efeitos aleatórios e pode ser derivada de tantas maneiras diferentes que tem o potencial de unificar as teorias estatísticas de modelos lineares. A presença de bons estimadores para os efeitos fixos e componentes da variância torna possível combinar informações de diferentes testes por algumas simplificações do BLUP. Este trabalho exemplifica as vantagens do Melhor Preditor Linear (BLP) aplicado ao melhoramento de plantas. Procurou-se ilustrar como proceder com estimativas de médias e de variâncias de progênie obtidas em diferentes ambientes para produzir preditores que tenham propriedades "melhor" (no sentido de variância mínima entre todas as combinações lineares entre as médias de progênie). Derivou-se uma regra prática para a produção dos pesos relativos de cada ambiente. O BLP, em alguns casos, é também não-viesado produzindo BLUPs a partir de lógica mais direta.

Termos para indexação: BLP, genética estatística, melhoramento de plantas.

(Received in august 23, 2006 and approved in june 16, 2008)

INTRODUCTION

One of the most typical features of plant genetical essays is the high level of balance and the better precision of variance component estimates compared with animal counterparts. Plant breeders do not always keep accurate records of genetic relatedness, in some allogamous species, for instance, open pollinated families are taken as half sib progenies for selection purposes. However, this is compensated by the large number of progeny and the replications of the trials in

same locations and years, conditions often impossible to attain with animal trials.

The usual analysis that guides plant breeders in selecting progenies tested in several environments is the pooled ANOVA, based on marginal averages of each family tested. The underlying assumption for this approach being that for each genetic value there is a constant effect. This is the heuristics of the fixed statistical modelling.

On the other hand, breeders know that the genetic values of individuals measured by the performance of their progenies are better viewed as random effects, representing

¹Doutor em Ciências Exatas, Professor Associado – Departamento de Ciências Exatas/DEX – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – juliobuenof@gmail.com

²Doutor, Professor Titular – Departamento de Genética – Escola Superior de Agricultura "Luiz de Queiroz"/ESALQ – Universidade de São Paulo/USP – Avenida Pádua Dias, 11 – Cx. P. 83 – 13400-970 – Piracicaba, SP – rvencovs@esalq.usp.br

small samples of the possible genotypes being tested. This way, the genetic value of a family in each environment can be viewed as one of the possible realizations of an unobservable random variable (the “true” breeding value). The intraclass correlation of those different realizations reflects the heritability for the selection based on progeny means. Thus, the appropriate form of analysis is more likely to follow the mixed model approach to progeny tests, which is a common practice in animal breeding after Henderson’s lifework.

In particular, Best Linear Unbiased Prediction (BLUP) has been considered as the most appropriate form in the analysis of genetic data in animal breeding trials. Following Robinson (1991) this BLUP is not a “method” but a feature of estimators (predictors) of random effects and can be derived in so many ways that has the potential of unifying the statistical theory of linear models.

However, the presence of fair estimates of fixed effects, coupled with a large amount of historical of data on variance components and heritability estimates makes it possible to combine information from several different tests by relaxing some BLUP assumptions. The purpose of this work is to introduce and exemplify the advantages of Best Linear Prediction (BLP) applied to plant breeding. Our focus is on how to deal with estimates of progeny means and variances from many environments to work out predictions that have “best” properties (in the sense of being minimum variance linear combinations).

In analogy with Robinson (1991), we think that if breeders can obtain good estimates of fixed effects, the BLP “method” will have also unbiased properties and we produce BLUPs from straightforward heuristics.

Statistical steps to establish BLUP as a reasonable classical predictor of genetic values may be found in some seminal papers since Henderson et al. (1959). In particular,

for BLP derivation and features like maximization of correct ranks (under normality assumptions), see Henderson (1963). For forestry breeding purposes, White & Hodge (1988) was the pioneer BLP work, and White & Hodge (1989) comprehensively covers both BLUP and BLP subjects. In forestry breeding, Resende et al. (1993) were the first Brazilian researchers to introduce BLP (and soon other mixed model techniques). Although these techniques are straightforward we have found no works in which BLP was applied to crop species.

METHODOLOGY

A progeny trial is a way to predict the breeding value of parents from realizations of its progeny phenotype. To make comprehensive selection decisions, plant breeders usually run the same trials in multiple environments (locations or years), Table I being a schematic representation of such trials.

The statistical model for each realization is:

$$\bar{y}_{ij} = m + A_j + p_i + (pA)_{ij} + \bar{e}_{ij}$$

In this model, m is the general mean, A_j is the fixed effect of environment j , p_i is the random effect of breeding value of progeny i ; $(pA)_{ij}$ is the random effect of interaction of i^{th} breeding value with j^{th} environment and \bar{e}_{ij} is the mean experimental error.

The underlying assumptions for BLP purposes are that the fixed effects and variance components of random effects of each level of the model are known. Then:

$$E(\bar{y}_{ij}) = m + A_j;$$

And the following variance component estimates are taken as true variance parameters:

Table 1 – Scheme for a multi-environment progeny test in which \bar{y}_{ij} are phenotypic values of phenotypic progeny means calculated for each environment.

	Progeny			Environment	
	A ₁	A ₂	...	A _J	Global means
P ₁	\bar{y}_{11}	\bar{y}_{12}	...	\bar{y}_{1J}	$\bar{y}_{1\bullet}$
P ₂	\bar{y}_{21}	\bar{y}_{22}	...	\bar{y}_{2J}	$\bar{y}_{2\bullet}$
.
.
P _I	\bar{y}_{I1}	\bar{y}_{I2}	...	\bar{y}_{IJ}	$\bar{y}_{I\bullet}$
Average	$\bar{y}_{\bullet 1}$	$\bar{y}_{\bullet 2}$...	$\bar{y}_{\bullet J}$	$\bar{y}_{\bullet\bullet}$

$\hat{\sigma}_p^2$: progeny variance; $\hat{\sigma}_{\bar{y}}^2$: phenotypic variance among overall means of progenies, a function of progeny and (average) error variances; $\hat{\sigma}_{\bar{y}_j}^2$: phenotypic variance among local means of progenies for environment j , that is a function of all the random term variances.

Note that this is a usual set of assumptions if the objective is selection. In this situation, all nuisance parameters are taken as fixed, as in animal breeding literature - for a comprehensive justification see White & Hodge (1988, 1989).

In this case BLP is calculated considering usual selection as a special case of BLP in which only one parameter (the global means in the right column of Table 1) is available for guiding the selection process.

At the next hierarchical level, with one progeny mean per environment (as displayed in Table 1) another set of breeding values with more general BLP properties may be calculated as weighted averages of environmental means. In this case, the weights are in some way inversely proportional to environmental (and non-additive genetical) variances.

When there are no population differences (fixed effects) among genetic values of progenies, the breeding values may be predicted by the following \hat{g} vector, that has BLP properties (SEARLE et al., 1992):

$$\hat{g} = E(g | y) = C'V^{-1}[y - E(y)]$$

in which C is the covariance matrix between genetic values and its phenotypic realization (t stands for a transposition operation); V^{-1} is the inverse of the covariance phenotypic data matrix; y is the phenotypic data vector and $E(.)$ stands for mathematical expectation.

Following White & Hodge (1989), both variance among the predictions and covariance between the predicted and true breeding values may be calculated by:

$$Var(\hat{g}) = Cov(\hat{g}, g) = C'V^{-1}C$$

A “goodness of fit” measure for the prediction process could be calculated by the correlation between the true and predicted genetic values:

$$r(\hat{g}, g) = \sqrt{\frac{Cov(\hat{g}, g)}{\sigma_g^2}}$$

in this expression σ_g^2 is the true genetic variance (assumed as known) of breeding values. This square root of the coefficient of genotypic determination is the so called “accuracy” and in our context will be used for comparison purposes.

RESULTS AND DISCUSSION

For the global means we can derive the result:

$$\hat{g}_i = h_m^2[\bar{y}_{ij} - \bar{y}_{..}]$$

Using all the elements in Table 1, for more than 2 environments, we get:

$$\hat{g}_i = b'[\bar{y}_{ij} - \bar{y}_{.j}]$$

in wich b is given by:

$$b = \frac{\sigma_p^2}{|V|} \begin{bmatrix} \prod_{j \neq 1}^J (\sigma_{\bar{y}_j}^2 - \sigma_p^2) \\ \prod_{j \neq 2}^J (\sigma_{\bar{y}_j}^2 - \sigma_p^2) \\ \cdot \\ \cdot \\ \cdot \\ \prod_{j=1}^{J-1} (\sigma_{\bar{y}_j}^2 - \sigma_p^2) \end{bmatrix}$$

in the above expression $|V|$ is the determinant of V matrix:

$$|V| = \det \begin{bmatrix} \sigma_{\bar{y}_1}^2 & \sigma_p^2 & \dots & \sigma_p^2 \\ \sigma_p^2 & \sigma_{\bar{y}_2}^2 & \dots & \sigma_p^2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_p^2 & \sigma_p^2 & \dots & \sigma_{\bar{y}_J}^2 \end{bmatrix}$$

Note that $\hat{\sigma}_{\bar{y}_i}^2$ includes progeny variance plus progeny-environment interaction variance as well as the environmental error variance estimate, although for only two environments it is not possible to separate these variance component estimates into distinct fractions.

Example

Let us take four environments with variance component estimates given by:

$$\hat{\sigma}_p^2 = 10 ; \hat{\sigma}_{\bar{y}_1}^2 = 15 ; \hat{\sigma}_{\bar{y}_2}^2 = 16 ; \hat{\sigma}_{\bar{y}_3}^2 = 18 \text{ and } \hat{\sigma}_{\bar{y}_4}^2 = 18.$$

The breeding value of *i* progeny can more realistic be calculated from:

$$\hat{g}_i = + 0.2791 [\bar{y}_{i1} - \bar{y}_{\bullet 1}] + 0.2326 [\bar{y}_{i2} - \bar{y}_{\bullet 2}] + 0.1744 [\bar{y}_{i3} - \bar{y}_{\bullet 3}] + 0.1744 [\bar{y}_{i4} - \bar{y}_{\bullet 4}].$$

The relative weights being 1.000, 0.833, 0.625 and 0.625.

The global mean approach in this case corresponds to rank genetic values with the same weight for each environment, regardless of the differences in precision of the estimated means. For the estimation of breeding progress by selecting the *ith* progeny, the deviance of the progeny mean must then be multiplied by the heritability of the selection based on global progeny means. The relative weights in this case are all 0.25.

As a practical rule - derived in Bueno Filho (1997), proofs shown in appendix - the relative weights of each environmental mean are in fact products of the differences between the phenotypic variances and the common progeny variance for the other environments.

In the example described in Table 1 we get the following weights (Table 2):

Plant breeders are often interested in selecting progenies tested in different locations and years, but in general each progeny is not tested in all environments.

Selection based on average environmental conditions is the most common objective, although in some special cases "target environments" are elected either by being the most (or least) productive locations, or by some desired (or undesired) climate, soil features, typical years with particular experimental properties, etc.

Although pooled ANOVA is a well-known technique, it is not designed to handle unbalanced data of this type which shows dependencies between factor levels. However, the application of more general BLP to such situations is straightforward, as highlighted in the example of Table 3, that shows experimental means from the

experiments described in Table 2, and Table 4 that shows the correspondent breeding values predictions.

The most important factor in determining relative weights for selection is the similarity of the progeny performance in both tested and target environment. For example, in Table 4 selection for A₁ uses the specific means of progenies in environment A₁ weighted by the progeny variance in environment A₁ (that includes progeny by environment interaction variance), while selecting for average environment this interaction is not included in the weighting process.

It is remarkable in Table 4 that progeny 3 has a positive predicted genetic value for environment A₁ that may lead it to be selected, although not tested! In conventional pooled ANOVA we do not even use the means of untested progeny and the target environment approach is unthinkable...

One of the most interesting features of best linear unbiased predictors (BLUP) is their ability to maximize the true ranks for normal populations. This is a theoretically proven fact which is well established by some 40 years of mixed model studies including Monte Carlo simulations. This means that rank differences between global means and BLP must lead to greater errors when using ANOVA, because the first, being more similar to BLUP, has less strong assumptions on the covariance structure. Examples that are sometimes given in which wrong ranking results from poor variance component estimates are insignificant for practical plant breeding purposes.

Registers on pedigree and molecular data information on genetic relatedness greatly increase the superiority of BLUP and BLP over pooled ANOVA.

Another very important fact about BLP (WHITE & HODGE, 1989) is that linear combinations of BLP have BLP properties. This suggests that to work out BLUP values rather than local means could lead to BLUP properties of breeding values, and these can be calculated in a BLP like

Table 2 – Relative weights for BLP selection based on environmental means of progenies:

Environmental Phenotypic Variance		Differences (Phenotypic-Progeny) Variances for remainder environments			Product <i>b./V</i>	Relative weight
A ₁ : $\hat{\sigma}_{\bar{y}_1}^2 = 15$		6	8	8	384	1.000
A ₂ : $\hat{\sigma}_{\bar{y}_2}^2 = 16$	5		8	8	320	0.833
A ₃ : $\hat{\sigma}_{\bar{y}_3}^2 = 18$	5	6		8	240	0.625
A ₄ : $\hat{\sigma}_{\bar{y}_4}^2 = 18$	5	6	8		240	0.625

Table 3 – Estimated means from 10 progenies in four environmental situations.

Progeny	Environment and Experimental Design				Global Mean
	A ₁ : simple lattice	A ₂ : RCB* with 4 replications	A ₃ : RCB with 3 replications	A ₄ : RCB with 3 replications	
1	5.6	7.2	6.2	6.1	6.29
2	5.8	7.1	6.1	6.1	6.26
3	-	7.1	6.0	5.4	6.16
4	5.9	7.2	6.0	5.5	6.16
5	5.9	6.9	6.0	5.6	6.11
6	5.9	6.6	-	5.8	6.10
7	5.4	6.9	5.9	5.9	6.03
8	5.2	7.4	5.9	5.6	6.02
9	5.0	6.8	6.1	-	5.99
10	5.3	6.7	6.6	5.4	5.97
<i>E(y)</i> (estimated by Environmental mean)	5.54	7.00	6.10	5.70	
Weights <i>b'</i> (target: average)	0.2791	0.2326	0.1744	0.1744	
Weights <i>b'</i> (target: E1**)	0.5674	0.1395	0.1047	0.1047	

- represent untested situations

* RCB is the "randomized complete blocks" experimental design

** let $\hat{\sigma}_{p1}^2 = \hat{\sigma}_p^2 + \hat{\sigma}_{pE1}^2 = 12$ be the variance of the progeny by environment interaction.

Table 4 – Estimated means from 10 progenies; predictions of breeding values for 2 targets of selection: average environment and selection for environment 1, with ranks for selection.

Progenies (ranked by global means)	Global Mean	BLP ₁ (average environment)	Rank by BLP ₁	BLP ₂ (target environment A1)	Rank by BLP ₂
1	6.29	0.16	1	0.08	1
2	6.26	0.15	2	0.05	2
3	6.16	-0.05	6	-0.05	8
4	6.16	0.10	3	0.01	4
5	6.11	0.05	4	-0.03	6
6	6.10	0.02	5	-0.69	10
7	6.03	-0.06	8	-0.01	5
8	6.02	-0.06	7	0.02	3
9	5.99	-0.18	10	-0.62	9
10	5.97	-0.13	9	-0.03	7

approach. This is as valid for multivariate analysis as for any other technique involving BLUP.

In the example given in this paper, it is possible to calculate BLUP values for all progenies and any target environment by using BLUP of progeny breeding values calculated independently in each environment. This potentially simplifies the computational task in calculating BLUP in a single model and allows simpler estimates that concatenates different years, locations, experimental designs, generations, types of progeny, relatedness of genetical material, etc.

Restricted Expectation of Maximum Likelihood (REML) is becoming a standard in likelihood-based analysis of linear mixed models (SEARLE et al., 1992). Although much of the previous BLP work demonstrates the readiness for its use at the operational level, statistical analysis of genetic trials is mainly concerned with producing complex models to manage REML like estimates and predictions directly.

However, a robust handling of progeny breeding values of progenies as random variables can be easily managed using BLP techniques from usual tables that

contains information from several environments in pooled analysis. To ignore these facts, results in greater estimation (prediction) errors and could increase progeny misclassification in selection trials.

AKNOWLEDGMENTS

The authors wish to thank Prof. Dr. Lucas Monteiro Chaves for insights on how to prove the matrix identities.

REFERENCES

BUENO FILHO, J. S. de S. **Modelos mistos na predição de valores genéticos aditivos em testes de progênes florestais**. 1997. 118 f. Tese (Doutorado) – Escola Superior de Agricultura de Luiz de Queiroz, Piracicaba, 1997.

HENDERSON, C. R. **Selection index and expected genetic advance**. Washington, DC: NRC, 1963. 982 p.

HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S.; KROSSIG, C. N. von. Estimation of environmental and

genetic trends from records subject to culling. **Biometrics**, Washington, v. 13, p. 192-218, 1959.

RESENDE, M. D. V. de; HIGA, A. R.; LAVORANTI, O. J. Predição de valores genéticos no melhoramento de *Eucalyptus*: melhor preditor linear. In: CONGRESSO FLORESTAL PAN-AMERICANO, 1., 1993, Curitiba, PR. **Anais...** Curitiba: IUFRO, 1993. p. 144-147.

ROBINSON, G. K. That BLUP is a good thing: the estimation of random effects. **Statistical Science**, v. 6, p. 15-51, 1991.

SEARLE, S.; CASELLA, G.; MCCULLOCH, C. E. **Variance components**. New York: J. Willey, 1992.

WHITE, T. L.; HODGE, G. R. Best linear prediction of breeding values in a forest tree improvement program. **Theoretical and Applied Genetics**, v. 76, p. 719-727, 1988.

WHITE, T. L.; HODGE, G. R. **Predicting breeding values with applications in forest tree improvement**. Dordrecht: Kluwer Academy, 1989.

APENDIX

Derivation of b'

We will consider the covariance matrices in a more amenable form without generality losses. This is achieved by taking a single vector of constants from C' and the following form for V :

$$c' = \mathbf{J}_{1 \times n} \text{ in which } \mathbf{J}_{n \times n} \text{ is the matrix that has only 1 as its elements;}$$

$$V_n = \mathbf{J}_n + \mathbf{D}_n, \text{ in which } \mathbf{D}_{ij} = d_i, \text{ if } i=j \text{ and } \mathbf{D}_{ij} = 0, \text{ if } i \neq j. \text{ Thus } d_i \text{ represents: } \frac{\sigma_{pAi}^2 - \sigma_p^2}{\sigma_p^2}$$

in which σ_{pAi}^2 can be any linear combination of variance components in which other terms plus additive genetical variance occurs (BUENO FILHO, 1997).

For any number (n) of environments, the determinant of V may be calculated as:

$$|V_n| = \prod_{i=1}^n d_i + \sum_{j=1}^n \prod_{\substack{i=1 \\ i \neq j}}^n d_i \quad (\text{equation 1})$$

This is a recurrencing relationship that may be proved by as follows. This relation is used to prove the analogous relation for the b vector, that has the following elements:

$$b_j = V_n^{-1} c = \frac{1}{|V_n|} \left[\prod_{\substack{i=1 \\ i \neq j}}^n d_i \right] \quad (\text{equation 2})$$

Equations (1) and (2) can easily be proved for special cases, e.g.:

$$|V_2| = \begin{vmatrix} 1+d_1 & 1 \\ 1 & 1+d_2 \end{vmatrix} = d_1 d_2 + d_1 + d_2 \quad b = \frac{1}{|V_2|} \begin{bmatrix} d_2 \\ d_1 \end{bmatrix}$$

$$|V_3| = \begin{vmatrix} 1+d_1 & 1 & 1 \\ 1 & 1+d_2 & 1 \\ 1 & 1 & 1+d_3 \end{vmatrix} = d_1 d_2 d_3 + d_1 d_2 + d_1 d_3 + d_2 d_3 \quad b = \frac{1}{|V_3|} \begin{bmatrix} d_2 d_3 \\ d_1 d_3 \\ d_1 d_2 \end{bmatrix}$$

Let call V_j^* the matrix that has element $d_j = 0$.

Taking equation (1) as true, it then follows that:

$$|V_k| = \prod_{i=1}^k d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i \quad (\text{equation 3})$$

and the same operations for including an environments in set k must be:

$$|V_{k+1}| = \prod_{i=1}^{k+1} d_i + \sum_{j=1}^{k+1} \prod_{\substack{i=1 \\ i \neq j}}^{k+1} d_i \quad (\text{equation 4})$$

Taking the last columns for calculating the determinants:

$$|V_n| = (1+d_n)(-1)^{2n} |V_{n-1}| + (-1)^{2n-1} |V_{2n-1}^*| + (-1)^{2n-2} |V_{2n-2}^*| + \dots + (-1)^{n+j} |V_j^*| + \dots + (-1)^{n+1} |V_1^*|$$

in which V_j^* is the nucleus of cofactors for row j and column n of V_n matrix. These matrices may be obtained by letting $d_j = 0$ as follows for V_2 and V_3 :

$$|V_{1(2)}^*| = 1$$

$$|V_{1(3)}^*| = \begin{vmatrix} 1 & 1+d_2 \\ 1 & 1 \end{vmatrix} = -d_2 \quad \text{and} \quad |V_{2(3)}^*| = \begin{vmatrix} 1+d_1 & 1 \\ 1 & 1 \end{vmatrix} = d_1$$

It may be easily shown that for odd j values this expression will be negative and positive for even j values. So for any j this will be:

$$|V_{n-1}^*| = (-1)^{2n-1} \begin{vmatrix} V_{n-2} & J_{n-1,1} \\ J_{1,n-1} & 1 \end{vmatrix} = -\prod_{i=1}^{n-2} d_i = -\prod_{\substack{i=1 \\ i \neq n-1}}^{n-1} d_i$$

$$|V_{n-2}^*| = (-1)^{2n-1} \begin{vmatrix} V_{n-3} & J_{n-3,1} & J_{n-3,1} \\ J_{1,n-3} & 1 & 1 \\ J_{1,n-3} & 1 & 1+d_{n-1} \end{vmatrix} = -\prod_{\substack{i=1 \\ i \neq n-2}}^{n-1} d_i \quad \text{and so on...}$$

The negative signals result from combination of odd exponents of cofactors or by changing the assortment of column $J_{n-1,1}$ when taking determinants of minor order. So, this expression can be simplified to:

$$|V_j^*| = (-1)^{2n-1} |V_j^*| = -\prod_{i=1}^{n-2} d_i = -\prod_{\substack{i=1 \\ i \neq n-1}}^{n-1} d_i \quad (\text{equation 6})$$

and V_{k+1} determinant result in:

$$|V_{k+1}| = (1+d_{k+1})(-1)^{2k+2} |V_k| - \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i \Rightarrow$$

$$|V_{k+1}| = (1+d_{k+1}) \left(\prod_{i=1}^k d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i \right) - \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i \Rightarrow$$

$$|V_{k+1}| = \prod_{i=1}^k d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i + d_{k+1} \left(\prod_{i=1}^k d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i \right) - \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i \Rightarrow$$

$$|V_{k+1}| = \prod_{i=1}^k d_i + \prod_{i=1}^{k+1} d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^{k+1} d_i \Rightarrow$$

$$|V_{k+1}| = \prod_{i=1}^{k+1} d_i + \prod_{i=1}^k d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^{k+1} d_i \Rightarrow$$

$$|V_{k+1}| = \prod_{i=1}^{k+1} d_i + \sum_{j=1}^{k+1} \prod_{\substack{i=1 \\ i \neq j}}^{k+1} d_i$$

, that is the same of equation (1) *q.e.d.*

In an analogous way, \mathbf{b}' in equation (2) is calculated by sums of the cofactors for \mathbf{j} columns as follows:

$$\begin{aligned} b'_{k+1} &= \frac{1}{|V_{k+1}|} \left[|V_k| + \sum_{j=1}^k (-1)^{n+j} |V_j^*| \right] \Rightarrow \\ b'_{k+1} &= \frac{1}{|V_{k+1}|} \left(\prod_{i=1}^k d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i - \sum_{j=1}^k (-1)^{n+j} |V_j^*| \right) \Rightarrow \\ b'_{k+1} &= \frac{1}{|V_{k+1}|} \left(\prod_{i=1}^k d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i + \sum_{j=1}^k (-1)^{n+j} |V_j^*| \right) \Rightarrow \\ b'_{k+1} &= \frac{1}{|V_{k+1}|} \left(\prod_{i=1}^k d_i + \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i - \sum_{j=1}^k \prod_{\substack{i=1 \\ i \neq j}}^k d_i \right) \Rightarrow \\ b'_{k+1} &= \frac{1}{|V_{k+1}|} \left(\prod_{i=1}^k d_i \right) = \frac{1}{|V_{k+1}|} \left(\prod_{\substack{i=1 \\ i \neq k+1}}^{k+1} d_i \right) \end{aligned}$$

and by analogy, we get for the following \mathbf{j} element:

$$b'_j = \frac{1}{|V_{k+1}|} \left(\prod_{\substack{i=1 \\ i \neq j}}^{k+1} d_i \right)$$

The target environment approach could be easily adapted by taking $1+d_i$ instead of 1 in the i element of the \mathbf{c}' covariance vector. This leads to an additional factor of $d_i b'_j$ in the weighting for environment \mathbf{j} .