Cantet, R., Kambo, S., Rodríguez, D. (2026) Ancestral regression to parents for the genetic evaluation of beef cattle, Proceedings 13th World Congress on Genetics Applied to Livestock Production 24283: https://doi.org/10.31274/wcgalp.24283
© 2026 *Cantet, et al.* CC BY-NC-ND

Ancestral regression to parents for the genetic evaluation of beef cattle

R.J.C. Cantet[1*], S.C. Kambo[2], and D. Rodríguez[3]

[1]Instituto de Investigaciones en Producción Animal, INPA-CONICET, Buenos Aires, Argentina; [2]Asociación Argentina de Brangus, Buenos Aires, Argentina; [3]Asociación Braford Argentina, Buenos Aires, Argentina [*]Corresponding author: rcantet@agro.uba.ar

Introduction

The prediction of breeding values (BV) of the Brangus and Braford breeds in Argentina have been continuously performed since 1999 (ERBra) and 2006 (PEGBraf), respectively. Brangus in 2020 and Braford in 2022 adopted the "ancestral regression to parents" (PAR) model of Cantet et al (2022) that adds the information of genomic arrays to phenotypes and pedigrees to refine genetic relationships between relatives. Our goal here is to present the results of six years of experience in running the PAR by the Animal Breeding Unit, at Universidad de Buenos Aires – INPA CONICET, for both breeds.

Materials & Methods

The AR and PAR models of genetic evaluation

Ideally pedigree and genomic information should be combined such that the covariance between relatives for BV allows for a Markov dependent structure. This is achieved by recursion of the BV of an animal in its parents, i.e. going backwards just one generation. This enabled the simple inversion of **A** (Henderson, 1975) that was instrumental of the success of the "animal model" (AM) of Henderson and Quaas (1976). As Cantet et al (2022, expressions (2) to (5)), the genetic relationships between relatives in AR and in PAR are multilocus probabilities of identity by descent" (IBD). Animal breeders have used one-locus probabilities of IBD between animals X and Y, i.e. $P(X \equiv Y)$ to define the elements of matrix **A**. To apply the notion of IBD to the genetic process of recombination, Weir and Cockerham (1969 a,b) defined two-locus probabilities of IBD $P(X_1 \equiv Y_1, X_2 \equiv Y_2)$. Eventually, Donnelly (1983) introduced continuous IBD to account for transmission of whole DNA segments. Therefore, for estimation purposes we employed these continuous IBD probabilities to parametrize the covariance structures the models presented below.

Cantet et al (2017) generalized the AM by combining pedigree and genomic information *to refine relatedness among animals* into an "ancestral regression" (AR). Let a_X be the BV of calf X. The BV of the ancestors of X are those of its sire (S), its dam (D), and those of the grandparents of X: the paternal grand-sire (SS), the paternal grand-dam (DS), the maternal grand-sire, and the maternal grand-dam (DD), respectively. The recursive equation for a_X in AR is then equal to:

$$a_X = 0.5\,a_S + 0.5\,a_D + \beta_S\left(a_{SS} - a_{DS}\right) + \beta_D\left(a_{SD} - a_{DD}\right) + \varphi_X \qquad (1)$$

The path coefficients for the paternal (β_S) and maternal (β_D) grandparents quantify the differential proportion of each grandparent in the genome of their grand-offspring that is produced during asymmetric recombination under meiotic interference, to form the parental gametes that will result in the genome of calf X. The residual BV in AR is φ_X. The parameters β_S and β_D are fully identifiable using genomic markers, such as dense SNP arrays. Cantet et al (2022, (11) page 4) provided the expression for the coefficient of inbreeding in AR.

Unfortunately, AR is a recursive model of inheritance but not Markov, because the recursion [1] depends on more than one ancestral generation. Due to this two-generation recursion the coefficient matrix in the mixed model equations (MME) requires up to six times more non-

zero elements for the genotyped animals than $\mathbf{A}^{-1}$. This increase in the number of coefficients is due to the inclusion of the path coefficients β_S for SS, $-\beta_S$ for DS, β_D for SD and $-\beta_D$ for DD of each genotyped animal in the inverse covariance structure of BV. The resulting coefficient matrix of the MME becomes dense even for moderate numbers of genotyped animals and computations become much more expensive than in AM.

To decrease the cost of computation Cantet et al (2022) "absorbed" all four grandparental path coefficients into the parental ones such that the recursive equation for a_X results in:

$$a_X = \rho_S a_S + \rho_D a_D + \phi_X \qquad (2)$$

The parameters ρ_S and ρ_D are the path coefficients of the BV of the sire and the dam, respectively, and the parameter space is $0.50 \leq \rho_S, \rho_D \leq 1/\sqrt{2} \approx 0.707$. The minimum limit is the same value as the constant 0.50 of the AM. The upper limit result from replacing in the residual variance (expression (28) in Cantet et al, 2022) such that the contribution of any grandparent to the genome of its grand-progeny is not less than 0.125 and not greater than 0.375, limit values observed in the real evaluation of Brangus and Brafords. Note that the residual variance cannot be 0, or BV will not be a normal random variable. In general, ρ_S and ρ_D may be numerically different depending on the differential information on the BV passed during meiosis. The ρ_S and ρ_D parameters reflect the amount of information from the BV of the grandparents of X that is actually being passed by the sire and dam of X to a_X, instead of the expected proportion of 0.25 for each grandparent in the AM. he difference between ρ_S or ρ_D with 0.50 is the amount of information that is being gained when predicting BV of X in comparison with the situation of an "idealized" transmission under random mating, free recombination, and neither linkage nor linkage disequilibrium.

Cantet et al (2022) observed that the absolute value of ρ - 0.5 reflects the difference between multilocus (*n*, say) and single locus probabilities of IBD: $P(X_1 \equiv Y_1, \ldots, X_n \equiv Y_n) - P(X \equiv Y)$. The quantity is a generalization of the identity disequilibrium (ID) concept of Weir and Cockerham (1969a,b). In theory, the higher the ID the larger the additive variance that may be available for prediction purposes, and the greater the true accuracy obtained.

If we set all BV in a vector **a**, the vector-matrix expression of [2] is $\mathbf{a} = \mathbf{B}\,\mathbf{a} + \varphi$, where **B** is the upper triangular matrix that contains ρ_S and ρ_D in the row of each progeny and in the columns corresponding to the BV of its parents. All remaining elements in the row are zero. Matrix **B** expresses the causal effect of parental BV in the BV of their progeny, because genomes and BV of parents always create or "cause" the genomes and BV of their progeny. Hence, it is impossible to have a ρ in the lower triangular part of **B** as it can occur in a correlational model. Residual BV in φ are independent random variables (Cantet et al, 2022) with a diagonal covariance structure **D**. Finally, the covariance structure of the BV in PAR ($\mathbf{\Sigma}$) and its inverse are respectively equal to

$$\mathbf{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1}\,\mathbf{D}\,(\mathbf{I} - \mathbf{B}')^{-1}; \quad \mathbf{\Sigma}^{-1} = (\mathbf{I} - \mathbf{B}')\,\mathbf{D}^{-1}\,(\mathbf{I} - \mathbf{B}) = \mathbf{D}^{-1} - \mathbf{B}'\mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{B} + \mathbf{B}'\mathbf{D}^{-1}\mathbf{B} \qquad (3)$$

The last expression to the right results in a similar algorithm to calculate $\mathbf{\Sigma}^{-1}$ without ever building $\mathbf{\Sigma}$ first, as Henderson (1975) did with $\mathbf{A}^{-1}$: just replace the 0.50 values for ρ_S and ρ_D into **B**. It can be proved that inbreeding in PAR for animal X is equal to half the PAR relationship between the parents of X, i.e. $F_{X\text{-}PAR} = 0.5\,\Sigma_{SD}$, which highlights the similitude with the AM since $F_X = 0.5\,\mathbf{A}_{SD}$.

Results

Comparison of PAR and GBLUP in MSEP. Harville (1985) decomposed mean squared error of prediction (MSEP) into four components : 1) inherent error; 2) unknown distributional form; 3) unknow means; 4) unknown variances and covariances. When comparing MSEP of PAR and GBLUP, we may assume that components 1, 2 and 3 act equally between both methods and concentrate on 4. Notice that all elements in **G** and the ρ - parameters should be estimated

for BLUP prediction. Prasad and Rao (2006) observed that the greater the number of parameters have to be estimated, the greater will be the underestimation of true MSEP. For k genotyped animals there are 2k and k + k(k□1)/2 parameters in in the covariance matrix of BV, for PAR and GBLUP respectively. Consequently, the amount of statistical information present in the data should be divided linearly by 2k for PAR, and by 0.5 k^2 for GBLUP. For example, if k = 1000, then PAR has 2,000 parameters and GBLUP 495,000 parameters. Therefore, component 4 of MSEP will increase more in GBLUP than in PAR. In practice, the comparison of PAR and GBLUP is obscured by the fact that the covariance matrices of BV are assumed to be correct in either method, so that MSEP is underestimated as Prasad and Rao (2006) observed. Then, underestimation of MSEP will be greater for GBLUP than for PAR.

Comparison of PEV in PAR and AM: The comparison of accuracies (a function of prediction error variances, PEV) from PAR with AM while assuming that each model is "true" is displayed in Figure 2. Cantet (1997) discussed comparison of prediction of BV with or without a true model. The figure displays the accuracies from AM (horizontal axis) with those from PAR (vertical axis) for all Brangus selection candidates that entered into the 2020 genetic evaluation, which was the first use of PAR in a genetic evaluation. Table 1 displays the average accuracy in PAR vs AM for all growth traits. Data belong to 1838 animals with ρ_S and ρ_D. A significant gain in accuracy from using PAR was observed for all traits when compared to AM. Notice that in those traits where there are less phenotypic information (Maternal weaning weight, Weight at 18 months and Scrotal Circumference) the effect of using ρ_S and ρ_D showed the largest gain in accuracy.

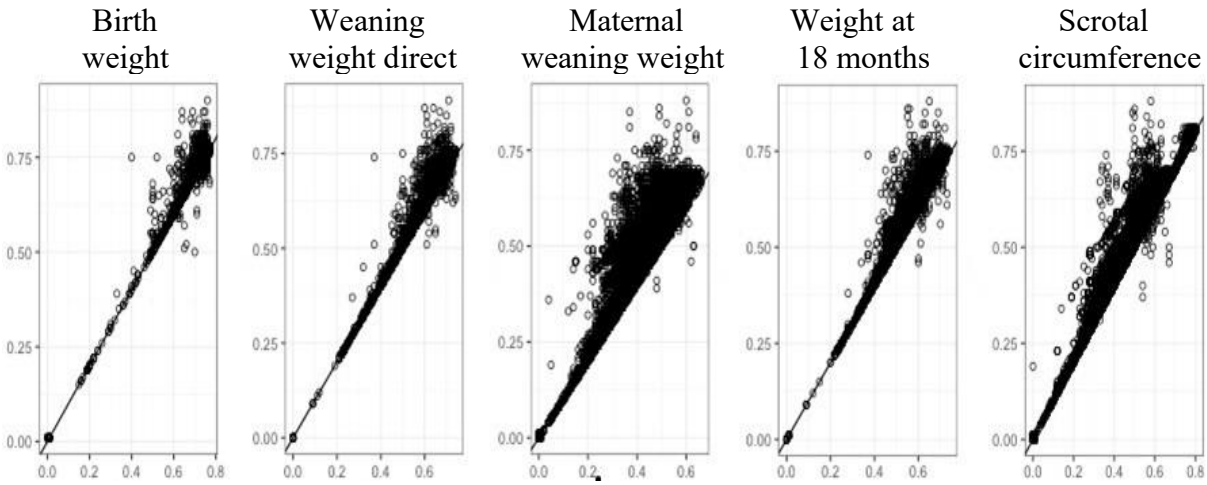

Figure 1. Comparison of accuracies from the "animal model" (AM, horizontal axis) with the ancestral regression to parents" (PAR, vertical axis).

Table 1. Averages accuracies from the "animal model" (AM) with the ancestral regression to parents" (PAR)

Mean accuracy	Birth Weight	Direct Weaning Weight	Maternal Weaning Weight	Weight at 18 months	Scrotal Circumference
PAR	0.72 ± 0.19	0.67 ± 0.20	0.58 ± 0.19	0.65 ± 0.19	0.62 ± 0.26
AM	0.69 ± 0.19	0.63 ± 0.20	0.46 ± 0.20	0.42 ± 0.19	0.55 ± 0.31

At the evaluation of December 2025, the multiple trait BLUP-PAR model for growth had more than three million equations in Brangus that were calculated in hour and a half, in a 64 GB desktop. The running time also includes the calculus of true accuracies, which are obtained with the efficient inversion algorithm of Takahashi et al (1973) for sparse symmetric matrices.

Discussion

We have recently submitted to a breeding journal the first of a series of papers that deal with the estimation of ρ_S and ρ_D, and the complexity of the problem precludes us from expanding on the topic here. Briefly, the estimating equations for ρ_S and ρ_D are solved using variants of the 2 - system of linear equations (24) in Cantet et al. (2022) with SNP arrays of seven individuals (SS,DS,SD,DD, S, D, and X) from three-generations.

An issue that proved to be challenging occurs when either the genotypes of SS or DS and those of SD or DD are missing, and even more difficult when either the sire or the dam are not genotyped. Notwithstanding this, the use of ancestral genotypic information through multilocus IBD relationships gives the PAR great flexibility to predict BV or the residual BV of animals that have records and lack genotypes of the relevant ancestors. Also, PAR predictions can be used to perform pure genomic selection on calves or young bulls and heifers that lack phenotypes, as well as with grade animals born from multiple-sire mating. The idea is to predict either the BV (or of the residual BV) with a selection index using PAR genetic relationships of the genotyped bulls and cows that make the largest genetic contributions to the evaluation. The first step is to estimate the genetic relationship of the animal with all relevant bulls and cows. To detect kinship, we test the genotype of the young animal to any relevant bull and cow in the pedigree of the breed with the allele frequency-free method of Waples et al (2018). Then, a minimal set of ancestors is chosen based on the largest and closest estimated kinships. Finally, the prediction of either the BV (or of the residual BV) of the animal is calculated with a selection index using the PAR predictions of those relevant ancestors.

Conclusions

The PAR is a recursive and Markov model that uses SNP arrays to refine relatedness among selection candidates. Predictions of BV from PAR have shown to be consistent across evaluations and years, with a sizeable increase in accuracy with respect to AM and with a similar computational effort. Prediction of BV of animals without phenotypic records, or without parents in the genetic evaluation, can be calculated using an IBD based procedure.

Acknowledgements

RJCC wants to thank to the Argentinian Brangus and Braford breeders that always supported the implementation into the ERBra and the PEGBRaf of the technical developments that resulted from research conducted at the University of Buenos Aires and INPA-CONICET.

References

Cantet R. J.C. (1997) J. Anim. Breed. Genet. 114:267-274.

Cantet R.J.C., Garcia-Baccino C.A., Rogberg-Muñoz A., Forneris N.S., Munilla S. (2017) J. Anim. Breed. Genet. 134:224–231

Cantet R.J.C., Angarita-Barajas B., Forneris N., Munilla S. (2022) Genet. Sel. Evol. 54,64.

Donnelly K.P. (1983) Theor. Pop. Biol.23:34-63.

Harville, D.A. (1985). J. Amer. Stats, Assoc. 80:132-138.

Henderson C.R. (1975) J. Dairy Sci. 58:1727-1730.

Henderson C.R., and Quaas R.L. (1976) J. Anim. Sci. 43:1188-1197.

Prasad, N.G.N. and Rao, J.N.K. 2006. J. Amer. Stats. Assoc. 85:163–171.

Takahashi K., Fagan J., Chin M.S. (1973) 8th PICA Conference Proceedings, Minneapolis, MN. USA.

Waples R.K., Albrechtsen A., Moltke. 2018. Mol. Ecol. 28:35-48.

Weir B.S., and Cockerham C.C. 1969a. 61:923–940

Weir B.S., and Cockerham C.C. 1969b. 63:711–742.