

Document downloaded from:

<http://hdl.handle.net/10251/169651>

This paper must be cited as:

Aguilar, I.; Fernandez, EN.; Blasco Mateu, A.; Ravagnolo, O.; Legarra, A. (2020). Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. *Journal of Animal Breeding and Genetics*. 137(4):356-364. <https://doi.org/10.1111/jbg.12470>



The final publication is available at

<https://doi.org/10.1111/jbg.12470>

Copyright Blackwell Publishing

Additional Information

Effects of ignoring inbreeding in model-based accuracy for

BLUP and SSGBLUP

Short title: ignoring inbreeding in model-based accuracy

Ignacio Aguilar¹, Eduardo N. Fernandez², Agustin Blasco³, Olga Ravagnolo¹, Andres Legarra⁴

¹ Instituto Nacional de Investigación Agropecuaria (INIA), Canelones, Uruguay

² Cátedra de Mejora y Conservación de Recursos Genéticos e Instituto de Investigación sobre Producción Agropecuaria, Ambiente y Salud, Facultad de Ciencias Agrarias, UNLZ, Buenos Aires, Argentina

³ Institute for Animal Science and Technology. Universitat Politècnica de València. P.O. Box 22012. València 46022. Spain

⁴ UMR GenPhySE, INRA Toulouse, BP52626, 31326 Castanet Tolosan. France

16

17

ABSTRACT

18 Model-based accuracy, defined as the theoretical correlation between true and estimated
19 breeding value, can be obtained for each individual as a function of its Prediction Error
20 Variance (PEV) and inbreeding coefficient F , in BLUP, GBLUP, and SSGBLUP genetic
21 evaluations. However, for computational convenience, inbreeding is often ignored in two
22 places. First, in the computation of $\text{reliability} = 1 - \text{PEV} / (1 + F)$. Second, in the set-up, using
23 Henderson's rules, of the inverse of the pedigree-based relationship matrix \mathbf{A} . Both
24 approximations have an effect in the computation of model based-accuracy and result in
25 wrong values. In this work, first we present a reminder of the theory and extend it to
26 SSGBLUP. Second, we quantify the error of ignoring inbreeding with real data in three
27 scenarios: BLUP evaluation and SSGBLUP in Uruguayan dairy cattle, and BLUP evaluations
28 in a line of rabbit closed for >40 generations with steady increase of inbreeding up to an
29 average of 0.30. We show that ignoring inbreeding in the set-up of the \mathbf{A} - inverse is
30 equivalent to assume that non-inbred animals are actually inbred. This results in an
31 increase of apparent PEV that is negligible for dairy cattle but considerable for rabbit.
32 Ignoring inbreeding in $\text{reliability} = 1 - \text{PEV} / (1 + F)$ leads to underestimation of reliability for
33 BLUP evaluations, and this underestimation is very large for rabbit. For SSGBLUP in dairy
34 cattle it leads to both underestimation and overestimation of reliability, both for genotyped
35 and non-genotyped animals. We strongly recommend to include inbreeding both in the set-
36 up of \mathbf{A} - inverse and in the computation of reliability from PEVs.

37

38 1. INTRODUCTION

39 The purpose of genetic evaluations is to predict, with some uncertainty, the breeding value
40 of animals. Model-based accuracy (or its square, reliability) is used as a measure of risk in
41 choosing parents of the next generation, and it condenses in a single number the
42 uncertainty related to its breeding value. The measure of accuracy from BLUP theory is
43 regularly used and reported in breeding evaluations (Misztal & Wiggans, 1988). The advent
44 of genomic selection needs methods to ascertain individual accuracies (Edel, Pimentel,
45 Erbe, Emmerling, & Götz, 2019), and the increasing selection for complex traits (e.g. feed
46 efficiency or methane emissions (Pryce et al., 2015)) needs measures of individuals
47 accuracies in small to medium size data sets. For historical reasons of simplicity,
48 inbreeding is often ignored in computations of accuracy from prediction error variance
49 (PEV). Furthermore, often, pedigree inbreeding is also ignored in the computation of the
50 inverse relationship matrix (**A**- inverse) using Henderson's (1976) rules, which results in
51 an approximated BLUP and a further level of approximation for the computation of model-
52 based individual accuracies. As a result, even for small data sets where exact computations
53 of PEVs are feasible, reported accuracies are often approximate.

54 The objective of this paper is to present correct ways of considering inbreeding in the
55 computation of accuracy and compare with results from ignoring inbreeding, in a single
56 step GBLUP and traditional pedigree-based BLUP context, using two datasets: a dairy cattle
57 data set with genotyped individuals (using BLUP and SSGBLUP) and a closed rabbit
58 selection line with a large number of generations and steady increase of inbreeding (using
59 BLUP).

60

61 2. MATERIAL AND METHODS

62 2.1 Theory

63 Accuracy (*acc*) is a model-based, individual measure of precision of the Estimated Breeding
64 Value (EBV). It is typically defined (e.g. VanVleck 1993a,b) as the correlation, on repeated
65 conceptual sampling, of the true breeding value of one individual (u) with its estimate (\hat{u}).
66 For BLUP models, and assuming that the model is true, Henderson (1975, 1982, 1984)
67 showed that $Cov(u, \hat{u}) = Var(\hat{u})$. Also, $Cov(u, \hat{u}) = Var(\hat{u}) = Var(u) - PEV$ where $PEV =$
68 $Var(\hat{u} - u)$ and can be numerically obtained from the corresponding element on the
69 inverse of the mixed model equations (either by sparse inversion or MonteCarlo methods).
70 We will work with reliability (*rel*), the square of accuracy. According to the definition of a
71 correlation:

$$72 \quad rel = acc^2 = \frac{Cov(u, \hat{u})^2}{Var(\hat{u})Var(u)} = \frac{Var(u) - PEV}{Var(u)} = 1 - \frac{PEV}{Var(u)}$$

73

74 It is often, but wrongly, assumed, even in textbooks (e.g. (Bijma, 2012; Misztal & Wiggans,
75 1988; Mrode & Thompson, 2005) that $Var(u) = \sigma_u^2$, the genetic variance, whereas $Var(u)$
76 depends on each individual. An example is inbred individuals who tend to be more
77 extreme. More precisely, for individual i , $Var(u_i) = A_{ii}\sigma_u^2$ if pedigree is used (classical
78 BLUP), $Var(u_i) = G_{ii}\sigma_u^2$ (GBLUP) and $Var(u_i) = H_{ii}\sigma_u^2$ (SSGBLUP), where A_{ii} , G_{ii} and H_{ii}
79 represent measures of self-relationships, i.e., they are equal to $(1 + F_i)$ where F_i is the

80 inbreeding coefficient . In BLUP and SSGBLUP pedigree- or genomic- based measures of
 81 inbreeding are used. Note that, in selected populations, there is a reduction in the genetic
 82 variance due to selection and Bulmer effect (Bijma, 2012) but this is never considered in
 83 reporting model-based reliabilities and we will not do so. With this proviso, the correct
 84 expression for the computation of model-based reliability, defined as the squared
 85 correlation between u and \hat{u} is

$$86 \quad rel_i = 1 - \frac{PEV_i}{(1 + F_i)\sigma_u^2}$$

87 where F_i is a measure of inbreeding. This derivation is implicit in Henderson's results but,
 88 to our knowledge, it was first explicitly published by (Van Vleck, 1993a, 1993b) although it
 89 has certainly been used before (Tier, Schneeberger, Hammond, & Fuchs, 1991) . In
 90 conventional pedigree evaluations, ignoring the term F_i in the denominator results in an
 91 underestimation of reliability for inbred animals in BLUP evaluations as F_i can only be
 92 positive.

93 However, for genomic evaluations by GBLUP, there is a higher variability of inbreeding, and
 94 even "negative" inbreeding is possible ($G_{ii} < 1$), which indicate higher heterozygosity than
 95 population average (Legarra, Lourenco, & Vitezica, 2018). As for SSGBLUP, genotyped
 96 animals have self-relationships G_{ii} , whereas non genotyped animals get an "improved"
 97 estimation of inbreeding in matrix \mathbf{H} of the form $H_{ii} - 1$, where $H_{ii} = A_{ii} +$
 98 $\mathbf{a}_{i,2}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{a}_{2,i}$ (Legarra, Aguilar, & Misztal, 2009) where $\mathbf{a}_{i,2} = \mathbf{a}'_{2,i}$ is the part of
 99 the i -th column of \mathbf{A} that corresponds to the genotyped individuals. Xiang et al., (2017)

100 used $rel = 1 - \frac{PEV}{(1+F_i)\sigma_u^2}$ with $F_i = H_{ii} - 1$ obtained from \mathbf{H} constructed for crossbreds using

101 metafounders, and they argued that ignoring this correction in this particular model with
102 metafounders resulted in spurious results.

103 Although there are several reports for small data sets of model-based reliabilities using
104 SSGBLUP, none uses the diagonal of \mathbf{H} . The extent of the effect of ignoring the diagonal of \mathbf{H}
105 in the computation of model-based reliabilities is unknown, and it can (wrongly) either
106 increase or decrease the reliabilities.

107 A second problem in computation of reliabilities is that frequently, Henderson's (1976)
108 simplified rules are used for computation of \mathbf{A}^{-1} (Golden, Brinks, & Bourdon, 1991;
109 Mehrabani-Yeganeh, Gibson, & Schaeffer, 2000; Van Vleck, 1993b), resulting in the use of
110 an approximate matrix, that we call \mathbf{A}^{*-1} , instead. However, inversion of \mathbf{A}^{*-1} to obtain a
111 relationship matrix of \mathbf{A}^* does not yield the expected results. Quoting Golden et al. (1991)
112 "Using an approximate \mathbf{A} -inverse is not the same as assuming an individual has covariance
113 to each of his parents of .5, and a diagonal element in \mathbf{A} of 1". The use of either \mathbf{A}^{*-1}
114 implicitly assumes an incorrect matrix of "true" relationships, that is $(\mathbf{A}^{*-1})^{-1} = \mathbf{A}^*$, when
115 we know that $\mathbf{A}^* \neq \mathbf{A}$. This results in incoherencies in the genetic model, different MME,
116 and therefore different PEV.

117 Based on our experience and some studies (Mehrabani-Yeganeh et al., 2000), in BLUP,
118 using \mathbf{A}^{*-1} instead of \mathbf{A}^{-1} seems to change very little the results of genetic evaluations
119 (EBVs) and is the default in the BLUPF90 family of programs (Misztal et al., 2002), in
120 MixBLUP (Ten Napel et al., 2017), and in PEST (Groeneveld, Kovac, & Wang, 1990).
121 Software that correctly computes \mathbf{A}^{-1} by default includes Wombat (Meyer, 2007) and
122 ASReml (Gilmour, Gogel, Cullis, Thompson, & Butler, 2009).

123 Accordingly, the default in SSGBLUP computations was to compute

$$124 \quad \mathbf{H}^{*-1} = \mathbf{A}^{*-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix}$$

125 where \mathbf{A}^{*-1} was computed *without* inbreeding in Henderson's rules but \mathbf{A}_{22}^{-1} was computed

126 inverting \mathbf{A}_{22} , that was computed *with* inbreeding in all cases. Again, the use of \mathbf{H}^{*-1}

127 (formed as $\mathbf{H}^{*-1} = \mathbf{A}^{*-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix}$) implicitly assumes an incorrect matrix of

128 "true" relationships, that is $(\mathbf{H}^{*-1})^{-1} = \mathbf{H}^* \neq \mathbf{H}$. This resulted in strong convergence

129 problems for large data sets which included inbred animals (Matilainen, Strandén, Aamand,

130 & Mäntysaari, 2018; Strandén, Matilainen, Aamand, & Mäntysaari, 2017). It is unclear how

131 ignoring inbreeding when computing \mathbf{A}^{*-1} or \mathbf{H}^{*-1} , affects the calculation of PEV

132 (computed by inversion) and therefore accuracies.

133 Putting all together, ignoring inbreeding affects computation of individual accuracies in

134 two different places: first, if the wrong denominator is used (in $rel_i = 1 - \frac{PEV_i}{(1+F_i)\sigma_u^2}$) and

135 second, if PEV is wrongly computed using either \mathbf{H}^{*-1} or \mathbf{A}^{*-1} . VanVleck (1993b) already

136 pointed out both problems, however, there are no published examples in livestock data

137 sets of the consequences of ignoring F_i in the reliability, and there is no description of this

138 problem specifically in GBLUP or SSGBLUP evaluations.

139

140 2.2 Datasets

141 The dairy cattle data set involves milk yield from the Uruguayan Holstein national genetic

142 evaluation. Dairy cattle records were provided by the Uruguayan National Dairy Herd

143 Improvement (MU – Mejoramiento y Control Lechero Uruguayo, Montevideo Uruguay) and
144 genealogical information was provided by the National Herdbook (ARU -Asociacion Rural del
145 Uruguay, Montevideo, Uruguay). Data consisted of 305 DIM milk yields from 925,821 records
146 of 377,612 cows from lactation 1 to 5 since 1990 to 2018. Pedigree file was created using 3
147 generations of ancestors backwards from either phenotyped or genotyped animals and consisted
148 of 511,576 animals. Figure 1 shows the average level of inbreeding for cows with phenotypes.
149 The model used was a single trait model with repeated records for milk yield which included
150 fixed effects of herd-year-season, lactation-age, and random effects of permanent environment
151 and animal additive genetic effect. This is not the official evaluation which is a random
152 regression model. Genetic groups were not used in this particular genetic evaluation, given that
153 the purpose was to compute reliabilities (which are not well defined under a fixed genetic group
154 model). Heritability of the trait was 0.20.

155 Genotypes of 5,072 animals were available for analyses, including 2,246 and 2,826
156 genotypes for bulls and cows respectively. Genotypes were obtained from the International
157 Dairy and Beef SNP chip IDBv3 (Mullen et al., 2013) by Weatherby's Scientific Ltd.
158 (Johnstown, Naas, Co. Kildare, Ireland) and from the Illumina Bovine SNP50k Beadchip
159 (Illumina Inc. San Diego, CA). The 39,288 SNP markers in common from both panels were
160 available for analyses. SNP with minor allele frequency greater of 0.05 from autosome
161 chromosomes and samples with a call rate >0.90 were used. Missing SNP were imputed
162 using FImpute (Sargolzaei et al., 2014) with the UMD 3.1 assembly. Distribution of animals
163 with phenotypes and genotypes is presented in Table 1. Genetic evaluation was carried out
164 using BLUPF90 (Misztal et al., 2002) , and for each scenario, prediction error variances
165 (PEV) were obtained from the inverse of MME using FSPAK-YAMS (Masuda, Aguilar,

166 Tsuruta, & Miształ, 2015). Therefore, different PEVs were obtained, according to each
167 assumed relationship matrix.

168

169 TABLE 1 HERE

170 FIGURE 1 HERE

171 The second dataset is from the meat rabbit line A from the Universitat Politècnica de
172 València, Spain. This is a closed selected line with complete records and pedigree since its
173 foundation. The breeding objective of this maternal line is litter size ($h^2 = 0.10$); further
174 description can be found in Fernandez et al. (2017). The data set for this work included
175 pedigree (40 generations totaling 5668 animals) and 15671 records (litter size) of animals
176 born from 1980 to 2009. Inbreeding increases steadily at a rate of 0.0084 per generation,
177 so that animals in 2009 have an average inbreeding coefficient of roughly 0.30. This data
178 set is included to verify the effects of high accumulated inbreeding on the computations of
179 accuracy. The linear model for genetic evaluation includes the fixed effects of inbreeding
180 depression (with F as covariate), parity-lactation status of the female, and random year-
181 season, additive genetic and permanent effects; see reference above for details. The high
182 cumulated inbreeding provokes that not fitting F as covariate in this data set yields biased
183 estimates of genetic trend (Fernandez et al. 2017).

184

185 *2.3 Models*

186 For the dairy cattle data, EBVs and accuracies were calculated using BLUP (only pedigree
187 information) and SSGBLUP (pedigree and marker information). Two scenarios for
188 computation of the inverse of the relationship matrix for the MME were tested:

- 189 • Correct A-inverse \mathbf{A}^{-1} (alternatively, H-inverse: \mathbf{H}^{-1}) using Henderson's rules with
190 inbreeding ("add_an_upginb" type of random effect in BLUPF90). This results in
191 correct *PEV*.
- 192
- 193 • Incorrect A-inverse \mathbf{A}^{*-1} (alternatively, H-inverse: \mathbf{H}^{*-1}) using Henderson's rules
194 ignoring inbreeding ("add_animal" type of random effect in BLUPF90). This results in
195 incorrect *PEV* that we will call *PEV**.

196 Matrix \mathbf{G} involved in \mathbf{H} was computed as described for \mathbf{G}_w in Christensen et al.
197 (2012)(Christensen, Madsen, Nielsen, Ostersen, & Su, 2012), i.e. using VanRaden's (2008)
198 method 1, adjusting for inbreeding and average relationship, and blending with 0.05 of
199 matrix \mathbf{A}_{22} .

200 In addition, BLUPF90 was modified to calculate reliabilities. Two computations of
201 reliabilities were done:

- 202 • Correct, accounting for inbreeding in the denominator $rel = 1 - \frac{PEV}{(1+F_i)\sigma_u^2}$, where F_i
203 may be from \mathbf{A} or from \mathbf{H} .
- 204 • Incorrect, *not* accounting for inbreeding in the denominator, $rel^* = 1 - \frac{PEV^*}{\sigma_u^2}$

205 For rabbit, since there was no marker information, only \mathbf{A}^{-1} or \mathbf{A}^{*-1} was computed for
206 BLUP, with resulting *PEV* or *PEV**, and accuracies were computed using as denominator

207 $(1 + F_i)\sigma_u^2$ or σ_u^2 . In particular, we computed the two reliabilities above (rel and rel^*) and
208 also $rel^\# = 1 - \frac{PEV^*}{(1+F_i)\sigma_u^2}$ which uses incorrect PEV and correct denominator.

209

210 3. RESULTS AND DISCUSSION

211 In both data sets, ignoring inbreeding in the setup of \mathbf{A}^{-1} or \mathbf{H}^{-1} resulted in negligible
212 changes in EBVs (correlations higher than 0.99).

213 *3.1 Dairy cattle: Effect of ignoring inbreeding in the relationship matrices*

214 When ignoring inbreeding in the setup of \mathbf{A}^{-1} , the (incorrectly built) inverses of the

215

216

217 numerator relationships are \mathbf{A}^{*-1} and \mathbf{H}^{*-1} . Use of these approximate matrices implies that
218 the (wrongly) assumed relationship matrices are $\mathbf{A}^* = (\mathbf{A}^{*-1})^{-1}$ and $\mathbf{H}^* = (\mathbf{H}^{*-1})^{-1}$. To
219 quantify the extent of the error of using \mathbf{A}^{*-1} and \mathbf{H}^{*-1} instead of \mathbf{A}^{-1} and \mathbf{H}^{-1} we plot the
220 diagonals of \mathbf{A} and \mathbf{H} (which contain correct estimates of inbreeding) versus the diagonals
221 of \mathbf{A}^* and \mathbf{H}^* (which contain incorrect estimates) (Figure 2). It can be seen that ignoring
222 inbreeding in Henderson's rules results in higher self-relationships, thus implicitly
223 assuming animals to be inbred when they in truth are non-inbred. This is in agreement
224 with Golden et al. (1991). Furthermore, inbred animals are considered more inbred than
225 they truly are. Correlations $r(\text{diag}(A), \text{diag}(A^*)) = 0.94$ and $r(\text{diag}(H), \text{diag}(H^*)) =$
226 0.87 indicate quite a strong modification in the assumptions of genetic composition.

227 For SSGBLUP, ignoring inbreeding in the setup of \mathbf{A}^{-1} , not only generates wrong diagonal
228 elements for \mathbf{A}^* , but also modifies diagonal elements for genotyped individuals in \mathbf{H}^* . In
229 addition, the block in $(\mathbf{H}^{*-1})^{-1}$ corresponding to genotyped individuals is no longer \mathbf{G} (the
230 genomic relationship matrix), but something different.

231

232 FIGURE 2 HERE

233

234 If an animal is (implicitly) assumed to be inbred by using simplified \mathbf{A}^{*-1} or \mathbf{H}^{*-1} , its *a*
235 *priori* uncertainty is higher, and this in turn results in an increase of the PEV computed
236 from the inverse of the MME (that we called PEV^*). Even though this is the case, the effect
237 of ignoring inbreeding on the numerical values of PEV^* was found to be small, i.e. $PEV^* -$
238 PEV tended to be negligible. For each extra 10% “implied wrong” inbreeding (as in Figure
239 2 for some animals with zero inbreeding) there is an extra increase of 0.5% in PEV. Figure
240 3 shows (correct) PEV vs. (incorrect) PEV^* . There is actually quite good agreement and
241 little dispersion. Correlation for PEV and PEV^* is greater than 0.99, and there is no
242 observable bias (i.e. PEV^* is on average neither larger nor smaller than PEV) and the
243 regression of PEV on PEV^* is very close to 1.

244

245 FIGURE 3 HERE

246

247 *3.2 Dairy cattle: Effect of ignoring inbreeding in the computation of the reliability*

248 In Figure 4 we present the results comparing the (worse possible) expression of
249 repeatability, $rel^* = 1 - \frac{PEV^*}{\sigma_u^2}$ with the (correct) $rel = 1 - \frac{PEV}{(1+F_i)\sigma_u^2}$. We choose not to
250 present intermediate cases, e.g. with incorrect numerator but correct denominator as this
251 makes presentation cumbersome. In the first case, inbreeding is ignored both in the
252 construction of \mathbf{A}^{-1} and \mathbf{H}^{-1} (and therefore in the MME) and in the computation of
253 reliability from PEV, which is the default option of many software such as BLUPF90 or
254 PEST. The second case is the theoretically sound option. It can be seen that ignoring
255 inbreeding systematically underestimates reliability in BLUP while in SSGBLUP there is
256 over and underestimation of reliability.

257 Another issue with ignoring F_i is that some animals will obtain a PEV^* greater than σ_u^2
258 having thus a negative reliability. This will also mean that for highly inbred animals with
259 little information (and therefore large PEV^*), accuracy will not be computed since the
260 square root of $rel^* = 1 - \frac{PEV^*}{\sigma_u^2}$ is not a real number for $rel^* < 0$.

261

262 FIGURE 4 HERE

263

264 *3.3 Rabbit*

265 The effect of ignoring inbreeding in the relationship matrices is dramatic, and shown in
266 Figure 5. In fact, the assumed inbreeding is much higher, which in turn leads to

267 underestimation of PEV^* as shown below. The overall correlation between PEV and PEV^*
268 is 0.92. For this dataset, given strong selection, we present average computed reliabilities
269 per generation. Figure 6 (top) shows stabilization of reliability ($rel = 1 - \frac{PEV}{(1+F_i)\sigma_u^2}$) after a
270 few generations. As the line becomes inbred, there is an increase in both $PEV = Var(u -$
271 $\hat{u})$ and an increase in $Var(u) = (1 + F)\sigma_u^2$. However, if PEV^* is computed (ignoring
272 inbreeding coefficients in the setup of \mathbf{A}^{-1}) then PEV^* increases too much, leading to
273 apparently decreasing reliabilities ($rel^\# = 1 - \frac{PEV^*}{(1+F_i)\sigma_u^2}$) with time (middle). This trend is
274 even more marked if reliabilities are computed using the wrong denominator, (i.e. as
275 $rel^* = 1 - \frac{PEV^*}{\sigma_u^2}$), in which case there are negative values of reliability, which are not
276 consistent with the theory (bottom). All in all, an analyst of this data would be puzzled
277 because the amount of information does not change with generations, yet reliability seems
278 to decrease if incorrect expressions are used.

279

280 FIGURE 5 HERE

281

282 FIGURE 6 HERE

283

284 4. DISCUSSION

285 In the dairy cattle data set, with low levels of inbreeding, there is little effect of inclusion (or
286 not) of inbreeding in the setup of \mathbf{A}^{-1} and \mathbf{H}^{-1} on the actual values of PEV . In the rabbit

287 data set, with high levels of inbreeding, ignoring inbreeding in the setup of \mathbf{A}^{-1} leads to
288 highly inflated values of PEV^* and therefore to too low figures for reliability. We therefore
289 strongly recommend to *always* include inbreeding in the setup of \mathbf{A}^{-1} and \mathbf{H}^{-1} , because not
290 doing it may result in larger underestimation of accuracy and its cost is negligible. The cost
291 of computing pedigree-based inbreeding (which is all that is needed for correct \mathbf{A}^{-1} and
292 \mathbf{H}^{-1}) is negligible compared to the overall cost of the genetic evaluation.

293 In both data sets, there are errors if the correct denominator $(1 + F_i)$ is not used. These
294 errors underestimate reliability in BLUP and under and over-estimate reliability in
295 SSGBLUP. The overestimation is because some animals are more heterozygote than the
296 average of the population ($H_{ii} < 1$), and therefore less variable *a priori*. The fact that some
297 animals have low apparent reliability may inhibit breeders from using them; for instance,
298 animals with low rel^* , say 0.40 (accuracy=0.63) may have actual rel around 0.55
299 (accuracy=0.75). The correct denominator must therefore be used. Computing F_i with
300 pedigree has negligible cost. However, computation of inbreeding in Single Step analysis
301 from H_{ii} mixing pedigree and genomic relationships is not obvious. For genotyped animals,
302 H_{ii} is simply G_{ii} computed from genotypes, which is straightforward as a cross-product on
303 the animal's genotype (VanRaden, 2008). For non-genotyped animals, $H_{ii} = A_{ii} +$
304 $\mathbf{a}_{i2}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{a}_{2i}$ (Legarra et al., 2009), this is a cumbersome expression to
305 compute. Other options include sparse inversion of \mathbf{H}^{-1} (as in this work and in Xiang et al.
306 (2017)) or indirect methods such as Colleau et al. (2017).

307 Neither of the data sets has high heritability. High heritability would result in smaller PEVs
308 and lower error in the different approximations that ignore inbreeding – for instance,
309 negative reliabilities would possibly not be observed.

310 All in all, the practical benefits of using correctly inbreeding in genetic evaluations and in
311 computation of reliabilities are the following. It compares animals fairly while an inbred
312 animal is penalized if inbreeding is ignored. It allows for finer selection decision (i.e.,
313 genomic “heterogeneity” is correctly accounted for). It avoids negative reliabilities, which
314 are puzzling to report. Last, using correctly inbreeding protects from unexpected problems,
315 such as lack of convergence due to not consideration of inbreeding in the \mathbf{H}^{-1} matrix in
316 SSGBLUP (Matilainen et al., 2018; Strandén et al., 2017).

317

318 DATA AVAILABILITY STATEMENT

319

320 Data in this work is used with the permission of (dairy cattle) the Uruguayan National
321 Dairy Herd Improvement (MU – Mejoramiento y Control Lechero Uruguayo, Montevideo
322 Uruguay) and genealogical information was provided by the National Herdbook (ARU -
323 Asociacion Rural del Uruguay, Montevideo, Uruguay), and (rabbit) from the Universidad
324 Politecnica de Valencia, and is not publicly available.

325

326

327 ACKNOWLEDGMENTS

328 Work supported by FEDER (Poctefa funds, project ARDI), INRA SelGen metaprogram
329 (project GDivSelGen), and project Lomas CYT FCA 22, Universidad Nacional de Lomas de
330 Zamora (Argentina). This project has received funding from the European Unions' Horizon
331 2020 Research & Innovation programme under grant agreement N°772787 -SMARTER. We
332 thank reviewers for insightful comments.

333

334

335

336

Category	Total	Genotyped	Ungenotyped
Cows	377,612	2,464	375,148
Sires	8,223	606	7,617
Other ancestors	124,089	350	123,739
Candidates to selection	1652	1,652	0

337 **Table 1. Distribution of animals across categories in the Uruguayan Holstein data set**

338

339

340

341

342 REFERENCES

343 Bijma, P. (2012). Accuracies of estimated breeding values from ordinary genetic
344 evaluations do not reflect the correlation between true and estimated breeding
345 values in selected populations. *Journal of Animal Breeding and Genetics*, 129(5), 345–
346 358.

347 Christensen, O., Madsen, P., Nielsen, B., Ostersen, T., & Su, G. (2012). Single-step methods
348 for genomic evaluation in pigs. *Animal*, 6, 1565–1571.

349 Colleau, J.-J., Palhière, I., Rodríguez-Ramilo, S. T., & Legarra, A. (2017). A fast indirect
350 method to compute functions of genomic relationships concerning genotyped and
351 ungenotyped individuals, for diversity management. *Genetics Selection Evolution*, *49*,
352 87. <https://doi.org/10.1186/s12711-017-0363-9>

353 E. N. Fernandez, J.P. Sanchez, R. Martinez, A. Legarra, M. Baselga. (2017). Role of inbreeding
354 depression, non-inbred dominance deviations and random year-season effect in
355 genetic trends for prolificacy in rabbit closed lines. *Journal of Animal Breeding and*
356 *Genetics*, *6*, 441–452.

357 Edel, C., Pimentel, E. C. G., Erbe, M., Emmerling, R., & Götz, K.-U. (2019). Short
358 communication: Calculating analytical reliabilities for single-step predictions.
359 *Journal of Dairy Science*, *102*, 3259–3265. <https://doi.org/10.3168/jds.2018-15707>

360 Gilmour, A. R., Gogel, B. J., Cullis, B. R., Thompson, R., & Butler, D. (2009). ASReml user guide
361 release 3.0. *VSN International Ltd, Hemel Hempstead, UK*.

362 Golden, B. L., Brinks, J. S., & Bourdon, R. M. (1991). A performance programmed method for
363 computing inbreeding coefficients from large data sets for use in mixed-model
364 analyses. *Journal of Animal Science*, *69*(9), 3564–3573.
365 <https://doi.org/10.2527/1991.6993564x>

366 Groeneveld, E., Kovac, M., & Wang, T. (1990). PEST, a general purpose BLUP package for
367 multivariate prediction and estimation. *Proceedings of the 4th World Congress on*
368 *Genetics Applied to Livestock Production, Edinburgh*, *13*, 488–491.

369 Henderson, C. R. (1976). A simple method for computing the inverse of a numerator
370 relationship matrix used in prediction of breeding values. *Biometrics*, *32*(1), 69–83.

371 Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and
372 genomic information. *J Dairy Sci*, 92(9), 4656–4663.

373 Legarra, A., Lourenco, D. A. L., & Vitezica, Z. G. (2018). *Bases for Genomic Prediction*.
374 Retrieved from <http://genoweb.toulouse.inra.fr/~alegarra/>

375 Masuda, Y., Aguilar, I., Tsuruta, S., & Misztal, I. (2015). Technical note: Acceleration of
376 sparse operations for average-information REML analyses with supernodal methods
377 and sparse-storage refinements. *Journal of Animal Science*, 93(10), 4670–4674.
378 <https://doi.org/10.2527/jas.2015-9395>

379 Matilainen, K., Strandén, I., Aamand, G. P., & Mäntysaari, E. A. (2018). Single step genomic
380 evaluation for female fertility in Nordic Red dairy cattle. *Journal of Animal Breeding
381 and Genetics*, 135(5), 337–348. <https://doi.org/10.1111/jbg.12353>

382 Mehrabani-Yeganeh, H., Gibson, J. P., & Schaeffer, L. r. (2000). Including coefficients of
383 inbreeding in BLUP evaluation and its effect on response to selection. *Journal of
384 Animal Breeding and Genetics*, 117(3), 145–151. [https://doi.org/10.1046/j.1439-
385 0388.2000.00241.x](https://doi.org/10.1046/j.1439-0388.2000.00241.x)

386 Meyer, K. (2007). WOMBAT—A tool for mixed model analyses in quantitative genetics by
387 restricted maximum likelihood (REML). *Journal of Zhejiang University. Science. B*,
388 8(11), 815–821. <https://doi.org/10.1631/jzus.2007.B0815>

389 Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., & Lee, D. H. (2002). BLUPF90 and
390 related programs (BGF90). *7th World Congress on Genetics Applied to Livestock
391 Production*, CD-ROM Communication N° 28-07. Montpellier, France.

392 Misztal, I., & Wiggans, G. R. (1988). Approximation of Prediction Error Variance in Large-
393 Scale Animal Models. *Journal of Dairy Science*, 71, 27–32.
394 [https://doi.org/10.1016/S0022-0302\(88\)79976-2](https://doi.org/10.1016/S0022-0302(88)79976-2)

395 Mrode, R., & Thompson, R. (2005). *Linear models for the prediction of animal breeding*
396 *values*. Cabi.

397 Mullen, M. P., McClure, M. C., Kearney, J. F., Waters, S. M., Weld, R., Flynn, P., ... Berry, D. P.
398 (2013). Development of a custom SNP chip for dairy and beef cattle breeding,
399 parentage and research. *Interbull Bulletin*, 47, 58–66.

400 Pryce, J. E., Gonzalez-Recio, O., Nieuwhof, G., Wales, W. J., Coffey, M. P., Hayes, B. J., &
401 Goddard, M. E. (2015). Hot topic: Definition and implementation of a breeding value
402 for feed efficiency in dairy cows. *Journal of Dairy Science*, 98(10), 7340–7350.
403 <https://doi.org/10.3168/jds.2015-9621>

404 Sargolzaei, M., J. Chesnais, and F. Schenkel. 2014. A new approach for efficient genotype
405 imputation using information from relatives. *BMC Genomics*, 15(1):478.
406 <https://doi.org/10.1186/1471-2164-15-478>

407 Strandén, I., Matilainen, K., Aamand, G. p., & Mäntysaari, E. a. (2017). Solving efficiently
408 large single-step genomic best linear unbiased prediction models. *Journal of Animal*
409 *Breeding and Genetics*, 134(3), 264–274. <https://doi.org/10.1111/jbg.12257>

410 Ten Napel, J., Vandenplas, J., Lidauer, M., Strandén, I., Taskinen, M., Mäntysaari, E., ...
411 Veerkamp, R. F. (2017). *MiXBLUP, user-friendly software for large genetic evaluation*
412 *systems–Manual V2*.

413 Tier, B., Schneeberger, M., Hammond, K., & Fuchs, W. C. (1991). Determining the accuracy of
414 estimated breeding values in multiple trait animal models. *Proceedings of the 9th*
415 *AAABG Conference*, 239–242.

416 Van Vleck, L. D. (1993a). *Selection index and introduction to mixed model methods*. Boca
417 Raton, Florida: CRC Press.

418 Van Vleck, L. D. (1993b). Variance of prediction error with mixed model equations when
419 relationships are ignored. *Theoretical and Applied Genetics*, 85(5), 545–549.
420 <https://doi.org/10.1007/BF00220912>

421 VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.*,
422 91(11), 4414–4423.

423 Xiang, T., Christensen, O. F., & Legarra, A. (2017). Technical note: Genomic evaluation for
424 crossbred performance in a single-step approach with metafounders. *Journal of*
425 *Animal Science*, 95(4), 1472–1480.

426

427

428

429 **Figure 1.** Average level of inbreeding rate for Uruguayan dairy cattle.

430 **Figure 2.** True self-relationships vs. implied self-relationships (using Henderson's rules
431 and ignoring inbreeding) for Uruguayan dairy cattle.

432 **Figure 3.** Correct PEV versus incorrect PEV* (using Henderson's rules and ignoring
433 inbreeding) for Uruguayan dairy cattle.

434 **Figure 4.** Correct reliability fully considering inbreeding (X-axis) or not (Y-axis) for
435 Uruguayan dairy cattle. Colors indicate candidates to selection, cows, and bulls.

436 **Figure 5.** True self-relationships vs. implied self-relationships (using Henderson's rules
437 and ignoring inbreeding) in the "A" rabbit line.

438 **Figure 6.** Exact (top), approximate using incorrect PEV^* and dividing by $1 + F$ (middle),
439 and approximate using incorrect PEV^* and *not* dividing by $1 + F$ (bottom), reliabilities as a
440 function of generation number in the "A" rabbit line.

441