R. Bras. Zootec., 54:e20240186, 2025
https://doi.org/10.37496/rbz5420240186

Breeding and genetics
Full-length research article

Revista Brasileira de Zootecnia

Brazilian Journal of Animal Science
e-ISSN 1806-9290
www.rbz.org.br

INTERGEN beyond Bayesian and genomics: Iteration on data and approximate accuracies capabilities for large-scale genetic analysis in animal and plant breeding

Vinícius Silva Junqueira[1]*, Gabriel Soares Campos[2], Fernando Flores Cardoso[3]

[1] Bayer Crop Science, Uberlândia, MG, Brasil.
[2] Department of Animal Biosciences, Interbull Centre, Uppsala, Sweden.
[3] Embrapa Pecuária Sul, Bagé, RS, Brasil.

***Corresponding author:**
junqueiravinicius@hotmail.com

Preprint deposit: March 22, 2024
https://doi.org/10.21203/rs.3.rs-4133359/v1

Received: November 5, 2024
Accepted: June 6, 2025

How to cite: Junqueira, V. S.; Campos, G. S. and Cardoso, F. F. 2025. INTERGEN beyond Bayesian and genomics: Iteration on data and approximate accuracies capabilities for large-scale genetic analysis in animal and plant breeding. Revista Brasileira de Zootecnia 54:e20240186.
https://doi.org/10.37496/rbz5420240186

Editors:
Lucas Lima Verardo
Ana Fabrícia Braga Magalhães

ABSTRACT - The INTERGEN package has been under constant development for the last two decades using the Fortran 90/95 programming language. The latest version of INTERGEN includes three different binaries: *intergen*, *intergeniod*, and *intergenacc. Intergen* has been consistently expanded to include the Best Linear Unbiased Prediction (BLUP) genomic module in recent years. The *intergeniod* and *intergenacc* binaries are new and aim to estimate genetic merit using iteration on data (IOD) and approximate accuracy. The present study aimed to provide a high-level overview of INTERGEN functionalities and to assess the reliability of the IOD genomic new module included in the *intergeniod*. Single-step GBLUP (ssGBLUP) models using beef cattle growth and conformation traits were used to estimate breeding values using *intergeniod* and compared with benchmarking software (*intergen*). The rank correlation for the breeding values between *intergeniod* and *intergen* was equal to the unit. The INTERGEN package can be applied to the genetic analysis of large-scale datasets using *intergen*, *intergeniod*, or *intergenacc*.

Keywords: BLUP, GEBV, multi-breed, parentage uncertainty, reaction norms

1. Introduction

Genome selection (GS) is the standard genetic evaluation approach in many animal and plant species, thanks to the availability and affordability of genotyping technologies (Misztal et al., 2020). Since the publication of the landmark paper (Meuwissen et al., 2001), countless studies have been conducted to apply various SNP-based models. As the number of genotyped individuals with varied molecular marker densities continues to grow, it has become imperative to develop new algorithms and computational solutions to manage the vast amounts of pedigree, genotypes, and phenotypes. In cattle breeding, the number of animals with SNP markers has increased significantly over the past decade, from just a few thousand in the mid to late 2000s to over seven million in 2024 U.S. Holsteins (https://queries.uscdcb.com/Genotype/cur_freq.html) (Weller et al., 2017; Wiggans et al., 2017). The swine and poultry industries are also realizing the benefits of GS (Ibáñez-Escriche et al., 2014; Samorè and Fontanesi, 2016). In the plant breeding industry, crops like maize were among the first to adopt GS, and now, virtually all crops worldwide are using these models (Chan et al., 2012; Nannas and Dawe, 2015; Zystro et al., 2021).

Over the years, several software packages have been developed to estimate breeding values and variance components using either frequentist or Bayesian methodologies (Misztal, 1990; Van Vleck and Cassady, 2004; Meyer, 2007). These packages aim to estimate the population (co)variance components and genetic parameters, which are needed for designing breeding strategies and assessing the genetic merit of animal and plant populations (Henderson, 1975). The mixed model equations (MME) method is commonly employed for this purpose, assuming that (co)variance components are known a priori for all random effects. In recent years, new methods have emerged to estimate breeding values with greater accuracy, and algorithms are being implemented to enhance efficiency in computing time and memory usage for large-scale genetic evaluations. Some of these methods involve constructing relationship matrices more efficiently (Henderson, 1976; VanRaden, 2008; Aguilar et al., 2010; Misztal, 2016).

INTERGEN is a software package utilized for genetic analysis, specifically aimed at estimating the genetic merit of animal and plant populations. The software is written in Fortran 90/95 and is free for research use. Although it is widely recognized in animal breeding, particularly for its Bayesian features, this is the first comprehensive scientific publication detailing the functionalities of INTERGEN for both animal and plant populations. This paper offers an overview of the INTERGEN computational package, primarily focusing on highlighting new features advantageous for large-scale genomic evaluations of these populations.

2. Material and methods

2.1. Details on the computational package

The INTERGEN package consists of a collection of routines written in Fortran 90/95 that are useful for animal and plant breeding genetic analysis. It includes different modules for managing sparse and dense matrix allocation and operations, MME solvers such as FSPAK and preconditioned conjugate gradient (PCG), Gibbs and Metropolis-Hastings sampling algorithms for variance component estimation, and Hash and IJA storage computational strategies (Misztal, 1999). Initially, it was developed to conduct Bayesian research involving complex structures, such as genotype-by-environment interactions utilizing reaction norm models with unknown covariates in multibreed populations with uncertain paternity. This encompasses considerations for heterogeneity of residual variances, robustness analysis, and reduced animal models (Cardoso and Tempelman, 2003; Cardoso and Tempelman, 2004). INTERGEN has since been updated to include genetic merit estimation, incorporating (or not) genomic information using BLUP, and normal priors for fixed effects (Sorensen and Gianola, 2007). More recently, various algorithms have been implemented to enhance the statistical modeling capabilities of the package, allowing it to handle diverse management situations commonly encountered in livestock, genomic information, iterative algorithms for solving MME without memory allocation, and approximate accuracy estimation. The package comprises three binary programs written in Fortran 90/95: *intergen*, *intergeniod*, and *intergenacc*. In this section, we will briefly describe the differences among the binaries. The following sections will provide additional details about the functionalities currently available in the package.

2.2. Intergen

The software *intergen* is the most comprehensive among the three options available for analyzing genetic data using Bayesian and frequentist methods. It employs computational techniques to manage sparse operations and solve the MME for single-trait, multi-trait, and random regression models, which are commonly used in genetic analysis. The software can estimate variance components (Bayesian only) and derive solutions. INTERGEN enables users to define complex models with unlimited random and fixed effects. The MME are constructed and allocated in memory for Bayesian and BLUP analysis.

The software shares most components (i.e., modules, subroutines, and functions) for building equations between the Bayesian and BLUP methods. The most computationally demanding step in the analysis involves executing matrix operations, including ordering, symbolic and numerical factorization, and inversion. Among these, factorization and sparse inversion operations are the most resource-intensive (Junqueira et al., 2022b). In BLUP analysis, the MME solutions can be derived by exact inversion of the coefficient matrix or by using iterative techniques like PCG. All operations in the software are executed by allocating the equations in memory. Therefore, the computational capacity of the software depends on the RAM available in the user's computer. Computational limitations arise in large-scale analyses when all equations cannot be allocated in memory, particularly in multi-trait genomic analyses, and when prediction error variance (PEV) is required from the coefficient matrix inverse. For example, when analyzing over 25,000 genotyped individuals in multi-trait models, the software *intergeniod* is necessary for solving mixed-model equations using iteration on data (IOD), as RAM would reach its limits on most servers.

2.3. Intergeniod

The *intergeniod* software utilizes the IOD technique with PCG to solve the MME and obtain solutions for all the fixed and random effects specified in the model. This software is recommended whenever BLUE and BLUP solutions for large datasets are required and allocating memory for all equations is not feasible due to memory constraints. Schaeffer and Kennedy (1986) were responsible for the first IOD implementation in animal breeding. These authors implemented IOD with successive overrelaxation (SOR), which requires sorting the equations in a predefined order to solve them efficiently when allocated in memory. However, it is not flexible enough to handle different models without modifying the algorithm (Lidauer et al., 1999; Strandén and Lidauer, 1999). On the other hand, PCG is generic enough to handle any model with unlimited covariates and cross-classified effects. Below is a generic version of the PCG solver algorithm. Assume an MME denoted as $\mathbf{Ax = b}$ in which $\mathbf{A}$ is the left-hand side matrix, $\mathbf{x}$ is the vector of unknown solutions, and $\mathbf{b}$ is the right-hand side matrix. In the PCG, some vectors need to be initialized before starting the loop process, which is supposed to stop when reaching the convergence criteria defined by the user (default is 10^{-12}). The initial values of the scalars and vectors are $n = 0$, $\mathbf{w} = \mathbf{0}$, $\mathbf{x} = \mathbf{0}$, $\mathbf{r}_0 = \mathbf{b} - \mathbf{Ax}_0$, $\mathbf{d}_0 = \mathbf{M}^{-1}\mathbf{r}_0$, $f_0 = \mathbf{r'}_0\mathbf{d}_0$.

Algorithm 1 Generic Preconditioned Conjugate Gradient Algorithm

1: **while** not converge **do**
2: n=n+1
3: $q_n = Ad_{n-1}$
4: $\alpha_n = f_{n-1}/d_n q_n$
5: $x_n = x_n + \alpha_n d_{n-1}$
6: **if** mod(n,50) = 0 **then**
7: $r_n = r_{n-1} - Ax_n$
8: **else**
9: $e_n = r_{n-1} - \alpha_n q_n$
10: $s_n = M^{-1} r_0$
11: $f_n = r_n s_n$
12: $\beta_n = f_n / f_{n-1}$
13: $d_n = s_n + \beta_n d_{n-1}$

Figure 1 - Description of a generic preconditioned conjugate gradient (PCG) algorithm implemented in the *intergeniod* software.

Note that only a few vectors must be allocated in memory in this solver. In each iteration, the algorithm updates the solution vector $\mathbf{x}_n$, the residual vector $\boldsymbol{r}_n$, and the search direction $\mathbf{d}_n$. The matrix-vector product $\mathbf{q}_n = \mathbf{A}\mathbf{d}_{n-1}$ is used to compute the step size (α_n). The f_n is a scalar representing the inner product of the residual and its preconditioned form. The direction vector is then updated using β_n, ensuring the conjugacy of successive directions. The most complex computation is related to derivation $\mathbf{q} = \mathbf{A}\mathbf{d}$ as it requires estimating all elements of the coefficient matrix without allocating them in memory and multiplying them by the vector $\mathbf{d}$. Tsuruta et al. (2001) stated that the preconditioner matrix ($\mathbf{M}$) influences the convergence rate of PCG. The more straightforward construction—and more computationally efficient in terms of memory savings—of this square and symmetric matrix can become a vector if only diagonals of the coefficient matrix are utilized. Other possibilities rely on block diagonal structure, resulting in faster convergence rates at the expense of using more memory.

2.4. Intergenacc

The *intergenacc* package is the third binary designed to estimate the approximate accuracies of estimated breeding values using the MME method. This software analyzes large-scale datasets in which accuracies are required, and dimensionality and memory limitations prevent inverting the coefficient matrix of MME. It estimates the accuracies of random effects, such as animal or maternal effects, adjusted for one fixed effect and one additional random effect. This correction method is consistent across algorithms. The method implemented has been detailed in several papers (Misztal and Wiggans, 1988; Misztal et al., 1991; Misztal et al., 1993; Misztal et al., 2013). The current version of *intergenacc* supports single-trait models, with or without genomic information, and can estimate additive and maternal accuracies in separate runs.

2.5. Definition of the model effects

As previously mentioned, the INTERGEN package offers various useful features for conducting genetic evaluations of animals and plants. This section presents a brief overview of these functionalities. For more detailed guidance on using the software, please refer to the manual or contact the research group directly.

The software allows five different types of effects to be used to build the MME matrices. These effects include cross (cross-classified), cov (covariate), unknowncov (cross-classified covariate for reaction norms via random regression), rnorm (reaction norms via random regression), and ram (reduced animal model). In the EFFECT section of the parameter file, these effects are listed along with the number of levels and any indication of a nested effect for random regression modeling. Additionally, in this section, users can indicate if they wish to save samples of each MCMC cycle in an external file during Bayesian analysis. The software can handle an unlimited number of effects.

2.6. Residual (co)variances

The *intergen* software accommodates six types of residual (co)variance structures. It can fit both single- and multi-trait models within Bayesian framework, and allows Gaussian and heavier-tailed alternatives like Student's *t* or Slash densities for estimating variance components. Users can specify three residual densities, allowing for either homoscedastic (i.e., homogeneous) or heteroskedastic error specifications. Below are further details on each of the allowed specifications:

Gaussian homoscedastic: normal distribution with homogeneous variance $\mathbf{e} \sim N(0, \sigma_e^2)$, with $\mathbf{e}$ as the vector of residual effects and σ_e^2 the residual variance common to all elements of MME. Currently, this is the only option allowed in the *intergeniod*.

Gaussian heteroskedastic: normal distribution with heterogeneous variance $\mathbf{e} \sim N(0, \sigma_{e(i)}^2)$, with $\mathbf{e}$ as the vector of residual effects and $\sigma_{e(i)}^2$ the residual variance defined for subclass *i.*

Student's *t* homoscedastic: the Student's *t* distribution (or simply *t* distribution) is a continuous probability distribution that generalizes the normal distribution with heavier tails with controlled parameters determined by *v* as the degree of freedom. This distribution was presented as an alternative for mitigating the effects of preferential treatment and deviant observations (Strandén and Gianola, 1998). Under the assumption of homogeneous residual variance distribution, it can be defined as $\mathbf{e} \sim p(e|\sigma_e^2, w_i) = N(0, \frac{\sigma_e^2}{w_i})$, in which $p(w_i|v_e)$ represents the lack of fit of the marginal density of **e** to the Gaussian distribution. Thus, the distribution of the phenotypes becomes $\mathbf{y} \sim t(0, v_e)$.

Student's *t* heteroskedastic: this distribution is similar to the Student's *t* homoscedastic, and the difference relies on the assumptions around the presence of heterogenous residual variance among the subclasses. Thus, we fit $\sigma_{e_i}^2$ in $\mathbf{e} \sim p(e_i|\sigma_{e_i}^2, w_i) = N(0, \frac{\sigma_{e_i}^2}{w_i})$.

Slash homoscedastic: extending the derivation of Student's *t* distribution, Slash arises when $p(w_i|v_e) = v_e w^{v_e - 1}$, in which $i = 1,2, \ldots, n$, $v_e > 0$, $0 < w_i < 1$ such that $\sigma_E^2 = \frac{v_e}{v_e - 1}\sigma_e^2$.

Slash heteroskedastic: in situations in which it is assumed heterogeneity of residual variances, $\mathbf{e} \sim p(e_i|\sigma_{e_i}^2, w_i) = N(0, \frac{\sigma_{e_i}^2}{w_i})$ with w_i fitted as Slash homoscedastic.

In the case of Student's *t* and Slash distributions, the Metropolis-Hastings algorithm is applied on every cycle of MCMC. Previous publications provide more details on the derivation of the heavy-tailed alternatives and their implications for prior, posterior, and marginal probabilities (Cardoso and Tempelman, 2003; Rosa et al., 2003; Cardoso and Tempelman, 2004; Cardoso et al., 2005; Cardoso et al., 2007).

2.7. Random effects

Intergen currently accepts ten different (co)variance matrices for the random effects. These matrices can be calculated internally or imported into the software. The (co)variance structures defined at this stage are then added to the (off-)diagonal elements of the coefficient matrix. BLUP and MCMC models currently accept the following random types: diagonal, add_sire, add_animal, add_an_upg, add_an_upginb, add_an_ms, add_an_mb, diag_mb, user_file, user_file_i, and norm_prior.

For Bayesian analysis, additional parameters are required to define the hyperparameters of the distributions correctly. A pedigree file is needed as input for all types except diagonal types (i.e., diagonal, diag_mb, and norm_prior). The random types add_an_ms and add_an_mb handle multiple sires and breeds, respectively. Further statistical details on model parameterization can be found in Cardoso and Tempelman (2003) and Cardoso and Tempelman (2004). Furthermore, these random types can fit genetic groups, assumed as fixed, by including dummy parents in the pedigree file. The multi-sire and breed inferences are only available in Bayesian analysis. In such cases, the user should define the hyperparameter values for prior distributions (Dirichlet, Inverse Gamma, or Wishart). The software will run defined cycles of Metropolis-Hastings for each MCMC iteration and adjust the variance of the prior distributions based on the acceptance rates during the burn-in phase of MCMC.

The norm_prior allows the inclusion of prior knowledge (i.e., μ and σ^2) on the fixed MME matrices of the model (Sorensen and Gianola, 2007). The value μ can be understood as the best estimate a priori, and σ^2 the variance associated with the uncertainty of that estimate. Thus, with lower values of σ^2 higher is the confidence of μ. In this scenario, the traditional mixed model equation is then adjusted as follows:

$$[\mathbf{C} + \mathbf{V}^{-1}]\mathbf{b} = [\mathbf{y} + \mathbf{V}^{-1}\mathbf{m}],$$

in which **C** is the coefficient matrix of MME, **b** is the vector of BLUE and BLUP solutions, $\mathbf{V}^{-1}$ is the (co) variance matrix for the fixed effects, and **m** is the vector of a priori means for each fixed effect level.

The MCMC and BLUP allow using three types of random effects: add_animal, add_an_upg, and add_an_upginb. Amongst these, add_animal is the simplest, and the other are extensions. This type

utilizes information from the pedigree file, including the individual, parent 1, and parent 2. Iteratively, the inverse of the numerator relationship matrix (**A**) is built using this information without considering the inbreeding coefficients. The method implemented utilizes the concepts described in Henderson (1976) to directly create the inverse of **A**. The **A** matrix has many nonzero elements for a well-structured pedigree, but its inverse has only a few nonzero elements, making its construction fast and computationally efficient following Henderson (1976) rules. The add_an_upg type is the animal model that builds the **A** matrix with unknown parent groups (Westell et al., 1988), and add_an_upginb is the animal model with unknown parent groups and inbreeding. These models also allow the inclusion of genomic information from an external file. More details about the genomic module will be presented in the next section. The software provides two types of random effects that offer more flexibility: user_file and user_file_i. Using these types, the user can create their own (co)variance matrices externally and then upload them to the software. This functionality allows users to use INTERGEN even when no built-in features are available.

2.8. Genomic module

Our in-house genomic module has been under development for over a decade, and we have continuously modified it to enhance its computational efficiency in handling genomic information. While most non-genomic analyses are managed efficiently using sparse formats (Misztal, 1990), genomic analysis demands more efficient programming. The genomic module includes several parallel operations utilizing MKL libraries (Intel, 2024) to optimize computational performance.

Depending on the content of the pedigree, this module can handle single-step GBLUP (Aguilar et al., 2010) or GBLUP (genomic BLUP) (VanRaden, 2008). Assume, for instance, the following model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e},$$

in which **y** is the vector of phenotypes; **X** and **Z** are incidence matrices relating levels to phenotypes of the fixed and random effects, respectively; **b** is the vector of solutions of fixed (or systematic) effects; and **a** is the vector of random effects solutions. The vector $\mathbf{a} \sim N(0, \mathbf{G}_0 \otimes \mathbf{H}^{-1})$, $\mathbf{H}^{-1}$ is the inverse of the matrix describing the genetic relationships of the size of all individuals included in the pedigree file. This relationship matrix can be built assuming only pedigree in which $\mathbf{H}^{-1}$ becomes $\mathbf{A}^{-1}$, only genotypes in which $\mathbf{H}^{-1}$ becomes $\mathbf{G}^{-1}$ (i.e., GBLUP), or both information jointly, also known as single-step GBLUP (Aguilar et al., 2010). The $\mathbf{G}_0$ matrix contains the additive (co)variances and can have a scalar or a symmetric multidimensional matrix depending on whether it is a single- or multi-trait model. The $\mathbf{e} \sim N(0, \mathbf{I} \otimes \mathbf{R}^{-1})$, and **R** contains the residual (co)variances of the model, and the dimension also depends on the number of traits under evaluation. Then, based on the above statistical model, the MME that *intergen* and *intergeniod* handle can be described as follows:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}_0 \otimes \mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

The INTERGEN package has a genomic module shared among its three binaries. The current version of the module (version 1.0) can handle GBLUP and ssGBLUP models, depending on the pedigree provided by the user. If the pedigree file used in the analysis is empty (i.e., **A** and $\mathbf{A}_{22}$ are diagonals), the software will build a GBLUP model by creating a genomic matrix. However, if the pedigree file is not empty, the software builds a ssGBLUP model. The software constructs the genomic and pedigree-based matrices as described below:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(\alpha\mathbf{G} + \beta\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix},$$

in which τ, α, β, and ω are scaling factors used to improve matrices equivalences, with default values as 1.00, 0.95, 0.05, and 1.00, respectively (Misztal et al., 2010; Lourenco et al., 2014; Aguilar et al., 2020), and $\mathbf{A}_{22}$ is the pedigree-based relationship matrix constructed using the ancestral and genotyped individuals.

2.9. User interface

The INTERGEN package includes three binaries that can be executed on UNIX systems (Linux and OSX) via a terminal command line. Users must create a parameter file to specify the model parameters according to the software guidelines. Once executed, the software generates various intermediate and final files, which are saved in the directory the software is executed.

Before using the software, the user must ensure that the Math Kernel Library (MKL) is installed on their computer or server. The Intel compilers have been free since 2020 and can be downloaded from the Intel website (https://www.intel.com/content/www/us/en/developer/tools/oneapi/fortran-compiler-download.html).

Although all functionalities presented in this paper are operational in the INTERGEN package, we will not provide a comprehensive list of results since they have already been reported in several other publications (Santana et al., 2013; Ribeiro et al., 2015; Ribeiro et al., 2018; Junqueira et al., 2018; Junqueira et al., 2022a). Instead, in this publication, we will present the *intergeniod* results as the new INTERGEN package software.

The phenotypic, genotypic, and pedigree Angus data used in the analysis were kindly provided by the Programa de Melhoramento de Bovinos de Carne (PROMEBO) of the Associação Nacional de Criadores (ANC) "Herd-Book Collares". In the following sections, we will describe the data and the procedures used to prepare them for the analysis.

2.10. Phenotypes

The Angus animals used in this analysis were born between 1979 and 2022 and were primarily raised on pasture in the southern region of Brazil. Various traits were measured, such as weaning weight gain (WWG), weaning conformation (WC), weaning hair coat (WHC), post-weaning gain (PWG), yearling conformation (YC), yearling hair coat (YHC), scrotal circumference (SC), rib-eye area (REA) at yearling age, backfat thickness (BFT) at yearling, rump fat thickness (RFT) at yearling, and intramuscular fat (IMF) at yearling. For more information on each trait, please refer to Table 1.

Scores of WC and YC are visual measures of the volume of the carcass, considering body length and rib depth. Each animal was assigned a score between 1 and 5, with 5 being the highest expression of the trait and 1 being the lowest relative to its contemporary group (CG). The CG consisted of animals of the same sex, born in the same year and season, and raised on the same farm under the same management

Table 1 - Number of records (Phenotypes), number of genotyped animals with records (Geno_Pheno), number of contemporary groups (CG), and descriptive statistics (mean, minimum, maximum, and standard deviation [SD]) for each trait used in the analysis

Trait	Phenotypes	Geno_Pheno	CG	Mean	Minimum	Maximum	SD
WWG	291,538	9,602	16,938	140.98	20.50	410.00	40.06
WC	268,559	9,732	18,381	3.18	1.00	5.00	1.09
WHC	113,390	8,033	8,093	2.01	1.00	3.00	0.71
PWG	193,849	7,872	14,512	154.74	0.74	693.83	77.69
REA	33,626	5,943	4,326	59.51	15.06	129.70	18.03
BFT	34,421	5,948	4,250	2.97	0.10	23.60	1.63
RFT	30,140	5,940	3,963	3.49	0.10	24.60	2.24
IMF	26,344	5,847	3,430	3.05	0.30	10.28	1.15
YC	186,584	8,215	18,433	3.27	1.00	5.00	1.07
SC	49,937	4,466	4,352	34.53	18.00	50.00	3.71
YHC	91,797	7,647	9,252	1.81	1.00	3.00	0.69

WWG - weaning weight gain; WC - weaning conformation; WHC - weaning hair coat; PWG - post-weaning gain; REA - rib-eye area; BFT - backfat thickness at yearling; RFT - rump fat thickness at yearling; IMF - intramuscular fat at yearling; YC - yearling conformation; SC - scrotal circumference at yearling; YHC - yearling hair coat.

conditions. The phenotypes were collected on the same date, and CG with fewer than two observations were excluded from the analysis. Additionally, any observations that exceeded 3.5 standard deviations from their corresponding CG were removed.

Scores of WHC and YHC range from 1 to 3, with 1 indicating a short hair coat, 2 indicating a medium hair coat, and 3 indicating a long hair coat (Reimann et al., 2018). Traits REA (cm^2), BFT (mm), and IMF (%) were measured using an ultrasound device on the region between the 12th and 13th ribs, transversely over the *longissimus* muscle. The ultrasound measured RFT (mm) between the *gluteus medius* and *biceps femoris* muscles on the animal's rump.

2.11. Genotypes

A total of 12,637 animals were genotyped using sixteen different commercial SNP panels. Table S1 contains details about the number of SNP in each panel and the overlapping markers among them. Quality control of the genotypes was carried out using the R/SNPStats package (Clayton, 2023), which removed samples with genotyping call rates below 0.90, heterozygosity values three standard deviations above or below the observed mean, mismatched sex, and duplicate records. Only SNP mapped to autosomes with call rates greater than 0.98, minor allele frequencies greater than 0.03, and a probability of deviation from the Hardy-Weinberg equilibrium greater than 10^{-7} were considered for the analyses. Lastly, SNP with the highest minor allele frequencies were retained when observed in the same position or when genotypes were highly correlated (r > 0.98). Excluding SNP with high correlation from the analysis enhances numerical stability, boosts computational efficiency, and avoids overrepresentation of genomic areas.

After quality control, a joint imputation combined the SNP from the fourteen panels. Following the editing and joining of the SNP, 74,227 SNP and 12,637 samples were available for imputation. Missing genotypes were imputed using Fimpute software version 3 (Sargolzaei et al., 2011).

2.12. Statistical models

Seven multi-trait ssGBLUP models were analyzed, varying the random terms included. The complete ssGBLUP model can be described as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Pm} + \mathbf{Wc} + \mathbf{e},$$

in which **X, Z**, **P**, and **W** are the incidence matrices associating each level of fixed, additive, and maternal effects to observations. The **y** is the vector of observations, and the vector of BLUP solutions for additive and maternal effects are correlated in the form $\begin{bmatrix}\mathbf{a}\\ \mathbf{m}\end{bmatrix} \sim N(0, \mathbf{G}_0 \otimes \mathbf{H}^{-1})$, with $\mathbf{H}^{-1}$ as the relationship matrix constructed combining pedigree and genotypes (Aguilar et al., 2010), and the (co)variance matrix for this correlated effect defined as $\mathbf{G}_0 = \begin{bmatrix} \sigma^2_{a_1} & \sigma_{a_1a_2} & \sigma_{a_1m_2} & \sigma_{a_1m_2} \\ & \sigma^2_{a_2} & \sigma_{a_2m_1} & \sigma_{a_2m_2} \\ & & \sigma^2_{m_1} & \sigma_{m_1m_2} \\ sym & & & \sigma^2_{m_2} \end{bmatrix}$, the $\sigma^2_{a(m)_x}$ is the additive (maternal) variance for each trait x={1,2}, $\sigma_{a(m)_1a(m)_2}$ is the additive (maternal) covariance between traits and effects. The vector $\mathbf{c} \sim N\left(0, \mathbf{I} \otimes \begin{bmatrix}\sigma^2_{c_1} & 0\\ 0 & \sigma^2_{c_2}\end{bmatrix}\right)$ is associated with the maternal permanent environment. The residual term is described as $\mathbf{e} \sim N\left(0, \mathbf{I} \otimes \begin{bmatrix}\sigma^2_{e_1} & \sigma_{e_1e_2}\\ \sigma_{e_2e_1} & \sigma^2_{e_2}\end{bmatrix}\right)$ with $\sigma^2_{e_{1(2)}}$ and $\sigma_{e_1e_2}$ as the residual variance and covariances, respectively.

The effects associated with each trait and the variance components estimated by *intergen* are presented in Table 2. The analyses were carried out on *intergen* (benchmarking) and *intergeniod* using

PCG for deriving fixed and random solutions. All traits included additive effect as random. A random maternal effect was included for BW and WWG, while a random maternal permanent effect was included for BW, WWG, WC, and WHC. The fixed effects are CG, age of dam, and linear and quadratic covariates for animal age. The number of animals included in each analysis and the number of rounds to the convergence of each bivariate ssGBLUP model are presented in Table 3. The convergence criteria for analyzing conformation traits were set at 10^{-20}, while those for continuous records were established at 10^{-12}.

Table 2 - Description of the variance components and genetic parameters utilized in the single-step GBLUP analysis

Trait	Variance components					
	σ_a^2	σ_m^2	σ_c^2	σ_e^2	h^2	h_m^2
WWG	150.00	30.00	55.00	350.00	0.26	0.06
WC	0.112	-	0.027	0.613	0.15	-
WHC	0.076	-	0.063	0.238	0.20	-
PWG	110.00	-	-	726.00	0.13	-
REA	20.000	-	-	35.000	0.36	-
BFT	0.160	-	-	0.800	0.17	-
RFT	0.400	-	-	1.000	0.28	-
IMF	0.400	-	-	0.600	0.40	-
YC	0.149	-	-	0.631	0.19	-
SC	2.588	-	-	3.300	0.44	-
YHC	0.096	-	-	0.264	0.27	-

WWG - weaning weight gain; WC - weaning conformation; WHC - weaning hair coat; PWG - post-weaning gain; REA - rib-eye area at yearling; BFT - backfat thickness at yearling; RFT - rump fat thickness at yearling; IMF - intramuscular fat at yearling; YC - yearling conformation; SC - scrotal circumference at yearling; YHC - yearling hair coat; σ_a^2 - additive variance; σ_m^2 - maternal variance; σ_p^2 - permanent environment variance; σ_c^2 - maternal permanent environment variance; σ_e^2 - residual variance; h^2 - additive heritability; h_m^2 - maternal heritability.

Table 3 - Description of the pedigree size (Pedigree) and the number of rounds (Rounds) to reach convergence of iteration on data (IOD) via preconditioned conjugate gradient (PCG) for each of the single-step GBLUP models in *intergeniod* software (all analyses utilized 12,637 genotyped animals)

Analysis	Pedigree	Rounds
WC-YC	465,138	838
WWG-PWG	523,362	332
WWG-SC	521,230	113
WWG-REA	409,275	487
WWG-IMF	409,124	272
WWG-BFT-RFT	420,092	1,885
WHC-YHC	217,522	1,207

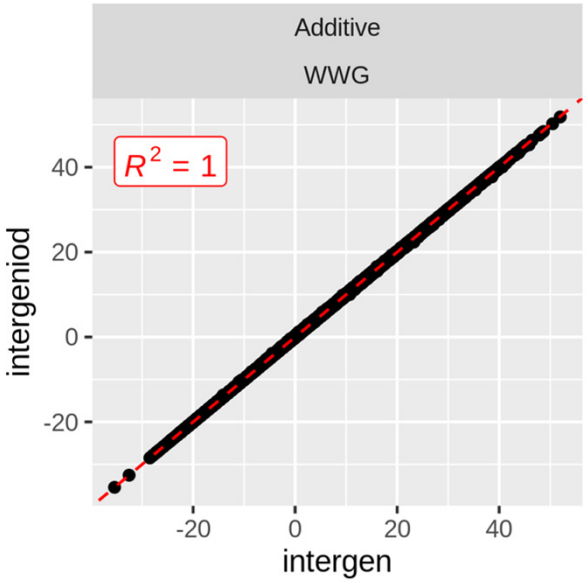
WC - weaning conformation; YC - yearling conformation; WWG - weaning weight gain; IMF - intramuscular fat at yearling; PWG - post-weaning gain; SC - scrotal circumference at yearling; REA - rib-eye area; WHC - weaning hair coat; YHC - yearling hair coat; BFT - backfat thickness at yearling age; RFT - rump fat thickness at yearling age.

2.13. Computing

All Fortran programs were compiled using an Intel Fortran compiler with maximum optimization. The analyses were performed on a Linux computer (x86-64) with an Intel® Xeon® Gold 5218R CPU running at 2.10GHz. To measure the random-access memory (RAM), the resident set size (RSS) and the virtual memory size (VSZ) displayed by the Linux ps command (https://man7.org/linux/man-pages/man1/ps.1.html) were summed up. The resulting value of ps is shown in kilobytes and then converted into gigabytes by dividing it by 1e6.

3. Results

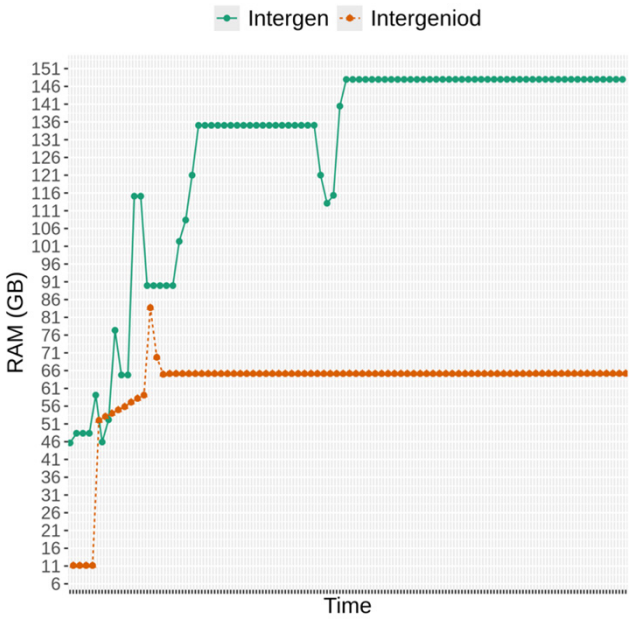
As mentioned earlier, only the results of the IOD will be presented in this publication, as the other functionalities are published elsewhere. Iteration on data is a technique that helps save RAM while solving MME in large-scale genetic analysis. If the specifications of the model are accurate and the convergence criterion is appropriate, the solutions produced by IOD should be similar to those obtained by storing equations in memory. The rank correlations between the solutions of all analyses are equivalent for all effects across all ssGBLUP models. Since all analyses had similar results, Figure 2 presents only the coefficient of determination (R^2) for the regression of the WWG additive solutions. The results of the additional traits are presented in Figure S1.



The goodness of fit of a linear regression of *intergen* on *intergeniod* solutions is presented as R^2.

Figure 2 - Scatter plot comparing the solutions for additive effect estimated using mixed model equations with iteration on data (*intergeniod*) and allocating equations in memory (*intergen*) with single-step GBLUP for weaning weight gain (WWG).

The RAM usage was monitored in all analyses, but only the WWG-BW bivariate ssGBLUP data results will be presented as the most extensive dataset. All other analyses showed similar results. As expected, *intergeniod* uses less than 50% of the RAM compared to *intergen* when the same model is allocated in memory (Figure 3). The current implementation of *intergeniod* allocates the **H** matrix and only a few vectors in memory. Due to genomic information, the **H** matrix consumes more RAM in IOD than any other structure. Therefore, enabling an algorithm in a future version of *intergeniod* that reads and processes the **H** matrix from a binary file saved in disk and stores it in buffer memory could save even more RAM. During the initial minutes of program execution, the genomic module allocates several intermediate matrices and vectors while constructing **H**, increasing RAM usage on both software. After the genomic module, the *intergeniod* enters a stable execution phase regarding RAM usage. On the other hand, *intergen* continues to increase RAM allocation as it executes matrix operations to derive the solutions.



Each dot represents 20 s.

Figure 3 - Description of the random-access memory (RAM) computer usage of *intergen* and *intergeniod* of a bivariate single-step GBLUP (ssGBLUP) model for a weaning weight gain and birth weight.

4. Discussion

This paper describes the functionalities of the INTERGEN package, which consists of three binaries for genetic studies of animal and plant populations. Our objective is to introduce the new capabilities of the package (iteration on data and approximate accuracy of genetic values) as Bayesian and genomic capabilities are well documented and published in the literature (Cardoso and Tempelman, 2003; Cardoso and Tempelman, 2004; Cardoso et al., 2005; Cardoso et al., 2007; Santana et al., 2013; Junqueira et al., 2018; Junqueira et al., 2022a).

Genomic selection has become the standard method for evaluating the genetic merit of various animal and plant species. Even when a species-specific marker panel is unavailable, advancements in second-generation sequencing technologies, such as RADseq (Andrews et al., 2016), allowed the identification of thousands of single-nucleotide polymorphisms without the need of a reference genome. This progress has expanded access to genomic applications (Fuentes-Utrilla et al., 2017; Marandel et al., 2020).

The advantages and disadvantages of utilizing genomic information are well established (Misztal et al., 2020; Misztal et al., 2021; Misztal and Lourenco, 2024). A significant challenge in applying genomic data in routine genetic evaluations is related to computational resources, particularly RAM, as the volume of data continues to grow (Junqueira et al., 2022b). There may be an ongoing need to develop more efficient algorithms for processor and memory utilization to manage this ever-increasing genotypic information. Numerous research groups are actively working on new algorithms that are beneficial for genetic analysis (Zaabza et al., 2023; Bermann et al., 2024; Ramos et al., 2024).

Iteration on data was initially introduced as an iterative method for constructing, computing, and solving mixed model equations (Schaeffer and Kennedy, 1986). It is a matrix-free iterative technique used in breeding value estimation. It solves mixed model equations by cycling through raw data records rather than explicitly constructing and allocating the full right- and left-hand side matrices in memory. In practice, an initial solution is iteratively refined by reading each observation, computing its residual,

and updating the relevant random and fixed effect estimates; this process repeats over all the data in each round until convergence.

Previously, the IOD technique was only needed for the largest national genetic evaluations before the genomic era. Now, it is a standard practice in most genetic evaluations to regularly incorporate genomic data to assess the genetic merit of individuals. Although initial implementations utilized successive overrelaxation techniques (Schaeffer and Kennedy, 1986), modern implementations often combine IOD with preconditioned conjugate gradient solvers to accelerate convergence on high-dimensional problems without appropriately sorting the equations. The implementation of IOD with PCG has been reported as highly efficient, flexible, and reliable in literature by various publications (Strandén and Lidauer, 1999; Tsuruta et al., 2001).

Iteration on data is widely applied—powering multi-trait animal models, random regression (longitudinal) analyses, and single-step genomic evaluations—due to its minimal memory requirements and ability to handle massive data by avoiding explicit cross-product computations. However, IOD does demand multiple passes through the dataset (read/write) and may converge slowly without strategic preconditioning. Therefore, recent variations have been focusing on improving efficiency (e.g., better preconditioners, parallelization, and hardware acceleration) to further scale genomic prediction in modern breeding programs.

As an iterative technique, it is important to ensure proper data structure and model effect specifications when solving MME with PCG. When considering the pedigree, it is advisable to remove unrelated individuals from genetic evaluations. In-house validations and personal communications with researchers worldwide have demonstrated that the convergence rate is significantly impacted, and sometimes, non-sense solutions are found if unrelated animals are kept in the pedigree file. It is also key to precisely define the effects of each trait when estimating MME solutions on IOD with PCG. Inappropriate model specifications can lead to local PCG convergence, especially in complex models such as multi-trait that incorporate genomic information (Pocrnic et al., 2017; Vandenplas et al., 2018).

4.1. Future implementations

The INTERGEN package is consistently updated to improve computational efficiency as the volume of genotype data increases. In the future, the package will include several new functionalities, such as estimating (co)variance components using restricted maximum likelihood (REML) (Harville, 1977) and average information REML (AI-REML) (Gilmour et al., 1995), robust modeling (Rosa et al., 2003; Cardoso et al., 2007), an algorithm for building a genomic matrix with proven and young (APY) strategy (Misztal, 2016), constructing the numerator relationship for selfing populations, and a multi-trait model for approximate accuracies (Strabel et al., 2001).

4.2. Research and commercial partnerships

The INTERGEN package has been used since 2016 by the PAMPAPLUS breeding program to estimate the genetic merit of beef cattle Hereford and Braford multibreed populations. In 2024, the PROMEBO breeding program adopted the *intergeniod* and *intergenacc* software to calculate the genetic merit of animals for eight different breeds. In these commercial partnerships, genetic evaluations are done each week, and the report that contains breeding values (additive and maternal) and BIF reliabilities is updated in the database for in-farm decisions.

5. Conclusions

The INTERGEN package is a software program written in the Fortran programming language. It consists of three binaries that can be used for the genetic analysis of large-scale datasets using Bayesian and BLUP analysis. The software *intergen* is a highly flexible tool and is free for research applications. Additionally, for situations in which the inverse of the coefficient matrix cannot be obtained directly

due to dimensionality issues, implementing IOD using *intergeniod* is a reliable alternative for estimating breeding values.

Supplementary material

The supplementary material of this article can be found online at: https://www.rbz.org.br/wp-content/uploads/articles_xml/1806-9290-rbz-54-e20240186/1806-9290-rbz-54-e20240186-suppl01.pdf

Data availability

The datasets for this article are not publicly available because they are the property of the Programa de Melhoramento de Bovinos de Carne (PROMEBO) producers and the Empresa Brasileira de Pesquisa Agropecuária (Embrapa), and this information is commercially sensitive. For scientific research purposes, the data requests should be forwarded along with the research proposal to F. F. Cardoso (fernando.cardoso@embrapa.br).

Author contributions

Conceptualization: Junqueira, V. S. and Cardoso, F. F. **Data curation:** Junqueira, V. S. **Formal analysis:** Junqueira, V. S. and Campos, G. S. **Funding acquisition:** Cardoso, F. F. **Investigation:** Junqueira, V. S. and Cardoso, F. F. **Methodology:** Junqueira, V. S. and Cardoso, F. F. **Resources:** Cardoso, F. F. **Software:** Junqueira, V. S. and Cardoso, F. F. **Supervision:** Junqueira, V. S. and Cardoso, F. F. **Validation:** Junqueira, V. S. and Campos, G. S. **Visualization:** Junqueira, V. S. and Campos, G. **Writing – original draft:** Junqueira, V. S. **Writing – review & editing:** Junqueira, V. S.; Campos, G. S. and Cardoso, F. F.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The authors thank Ignacy Misztal for providing the initial libraries utilized in the development of the INTERGEN package, Daniela Lourenço for her support and availability to discuss results during the validation of the IOD algorithm, and Yutaka Masuda for the discussions on linear algebra Fortran programming during the first author stay at the University of Georgia (Athens, GA, USA).

Financial support

The research was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Grant 481565/2004-3), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS - Grant 04/0808.4), and Embrapa (Grant 03.04.3.26).

References

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

Aguilar, I.; Misztal, I.; Johnson, D. L.; Legarra, A.; Tsuruta, S. and Lawlor, T. J. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93:743-752. https://doi.org/10.3168/jds.2009-2730

Andrews, K. R.; Good, J. M.; Miller, M. R.; Luikart, G. and Hohenlohe, P. A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17:81-92.

Bermann, M.; Aguilar, I.; Alvarez Munera, A.; Bauer, J.; Šplíchal, J.; Lourenco, D. and Misztal, I. 2024. Approximation of reliabilities for random-regression single-step genomic best linear unbiased predictor models. JDS Communications 5:582-586.

Cardoso, F.; Rosa, G. and Tempelman, R. 2007. Accounting for outliers and heteroskedasticity in multibreed genetic evaluations of postweaning gain of Nelore-Hereford cattle. Journal of Animal Science 85:909-918.

Cardoso, F. F.; Rosa, G. J. M. and Tempelman, R. J. 2005. Multiple-breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. Journal of Animal Science 83:1766-1779. https://doi.org/10.2527/2005.8381766x

Cardoso, F. F. and Tempelman, R. J. 2003. Bayesian inference on genetic merit under uncertain paternity. Genetics Selection Evolution 35:469. https://doi.org/10.1186/1297-9686-35-6-469

Cardoso, F. F. and Tempelman, R. J. 2004. Hierarchical Bayes multiple-breed inference with an application to genetic evaluation of a Nelore-Hereford population. Journal of Animal Science 82:1589-1601. https://doi.org/10.2527/2004.8261589x

Chan, C.; Qi, X.; Li, M.; Wong, F. and Lam, H. 2012. Recent developments of genomic research in soybean. Journal of Genetics and Genomics 39:317-324. https://doi.org/10.1016/j.jgg.2012.02.002

Fuentes-Utrilla, P.; Goswami, C.; Cottrell, J. E.; Pong-Wong, R.; Law, A.; A'Hara, S. W.; Lee, S. J. and Woolliams, J. A. 2017. QTL analysis and genomic selection using RADseq derived markers in Sitka spruce: the potential utility of within family data. Tree Genetics & Genomes 1-12.

Gilmour, A. R.; Thompson, R. and Cullis, B. R. 1995. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51:1440-1450. https://doi.org/10.2307/2533274

Harville, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 72:320-338. https://doi.org/10.2307/2286796

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-447. https://doi.org/10.2307/2529430

Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 69-83. https://doi.org/10.2307/2529339

Ibáñez-Escriche, N.; Forni, S.; Noguera, J. L. and Varona, L. 2014. Genomic information in pig breeding: Science meets industry needs. Livestock Science 94-100. https://doi.org/10.1016/j.livsci.2014.05.020

Junqueira, V. S.; Campos, G. S. and Cardoso, F. F. 2022a. INTERGEN: An efficient and flexible tool for large-scale genetic evaluation of complex animal and plant populations. p.1494. In: 12th World Congress on Genetics Applied to Livestock Production. Wageningen Academic Publishers, Rotterdam, Netherlands.

Junqueira, V. S.; Lopes, P. S.; Resende, M. D. V.; Silva, F. F.; Lourenço, D. A. L.; Yokoo, M. J. I. and Cardoso, F. F. 2018. Impact of embryo transfer phenotypic records on large-scale beef cattle genetic evaluations. Revista Brasileira de Zootecnia 47:e20170033. https://doi.org/10.1590/rbz4720170033

Junqueira, V. S.; Lourenco, D.; Masuda, Y.; Cardoso, F. F.; Lopes, P. S.; Silva, F. F. and Misztal, I. 2022b. Is single-step genomic REML with the algorithm for proven and young more computationally efficient when less generations of data are present? Journal of Animal Science 100:skac082. https://doi.org/10.1093/jas/skac082

Lidauer, M.; Strandén, I.; Mäntysaari, E. A.; Pösö, J. and Kettunen, A. 1999. Solving large test-day models by iteration on data and preconditioned conjugate gradient. Journal of Dairy Science 82:2788-2796. https://doi.org/10.3168/jds.S0022-0302(99)75536-0

Lourenco, D. A. L.; Misztal, I.; Tsuruta, S.; Aguilar, I.; Lawlor, T. J.; Forni, S. and Weller, J. I. 2014. Are evaluations on young genotyped animals benefiting from the past generations? Journal of Dairy Science 97:3930-3942. https://doi.org/10.3168/jds.2013-7769

Marandel, F.; Charrier, G.; Lamy, J. B.; Le Cam, S.; Lorance, P. and Trenkel, V. M. 2020. Estimating effective population size using RADseq: Effects of SNP selection and sample size. Ecology and Evolution 10:1929-1937.

Meuwissen, T. H. E.; Hayes, B. J. and Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829. https://doi.org/10.1093/genetics/157.4.1819

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

Misztal, I. 1990. Restricted maximum likelihood estimation of variance components in animal model using sparse matrix inversion and a supercomputer. Journal of Dairy Science 73:163-172. https://doi.org/10.3168/jds.S0022-0302(90)78660-2

Misztal, I. 1999. Complex models, more data: Simpler programming? p.33-33. In: Computational cattle breeding '99 workshop. Tuusula, Finland.

Misztal, I. 2016. Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics 202:401-409. https://doi.org/10.1534/genetics.115.182089

Misztal, I.; Aguilar, I.; Legarra, A. and Lawlor, T. J. 2010. Choice of parameters for single-step genomic evaluation for type. In: ASAS Annual Meeting.

Misztal, I.; Aguilar, I.; Lourenco, D.; Ma, L.; Steibel, J. P. and Toro, M. 2021. Emerging issues in genomic selection. Journal of Animal Science 99:skab092. https://doi.org/10.1093/jas/skab092

Misztal, I.; Lawlor, T. J.; Short, T. H. and Wiggans, G. R. 1991. Continuous genetic evaluation of Holsteins for type. Journal of Dairy Science 74:2001-2009. https://doi.org/10.3168/jds.S0022-0302(91)78369-0

Misztal, I.; Lawlor, T. J. and Short, T. H. 1993. Implementation of single- and multiple-trait animal models for genetic evaluation of Holstein type traits. Journal of Dairy Science 76:1421-1432. https://doi.org/10.3168/jds.S0022-0302(93)77473-1

Misztal, I. and Lourenco, D. 2024. Potential negative effects of genomic selection. Journal of Animal Science 102:skae155. https://doi.org/10.1093/jas/skae155

Misztal, I.; Lourenco, D. and Legarra, A. 2020. Current status of genomic evaluation. Journal of Animal Science 98:skaa101. https://doi.org/10.1093/jas/skaa101

Misztal, I.; Tsuruta, S.; Aguilar, I.; Legarra, A.; VanRaden, P. M. and Lawlor, T. J. 2013. Methods to approximate reliabilities in single-step genomic evaluation. Journal of Dairy Science 96:647-654. https://doi.org/10.3168/jds.2012-5656

Misztal, I. and Wiggans, G. R. 1988. Approximation of prediction error variance in large-scale animal models. Journal of Dairy Science 27-32. https://doi.org/10.1016/S0022-0302(88)79976-2

Nannas, N. J. and Dawe, R. K. 2015. Genetic and genomic toolbox of *Zea mays*. Genetics 199:655-669. https://doi.org/10.1534/genetics.114.165183

Pocrnic, I.; Lourenco, D. A. L.; Bradford, H. L.; Chen, C. Y. and Misztal, I. 2017. Impact of pedigree depth on convergence of single-step genomic BLUP in a purebred swine population. Journal of Animal Science 95:3391-3395. https://doi.org/10.2527/jas.2017.1581

Ramos, P.; Garcia, A.; Retallik, K.; Bermann, M.; Tsuruta, S.; Misztal, I.; Veroneze, R. and Lourenco, D. 2024. Comparing algorithms to approximate accuracies for single-step genomic best linear unbiased predictor. Journal of Animal Science 102:skae195. https://doi.org/10.1093/jas/skae195

Reimann, F. A.; Boligon, A. A.; Campos, G. S.; Cardoso, L. L.; Junqueira, V. S. and Cardoso, F. F. 2018. Genetic parameters and accuracy of traditional and genomic breeding values for eye pigmentation, hair coat and breed standard in Hereford and Braford cattle. Livestock Science 44-50. https://doi.org/10.1016/j.livsci.2018.04.007

Ribeiro, S.; Eler, J. P.; Pedrosa, V. B.; Rosa, G. J. M.; Ferraz, J. B. S. and Balieiro, J. C. C. 2015. Genotype × environment interaction for weaning weight in Nellore cattle using reaction norm analysis. Livestock Science 176:40-46. https://doi.org/10.1016/j.livsci.2015.03.032

Ribeiro, S.; Eler, J. P.; Pedrosa, V. B.; Rosa, G. J. M.; Ferraz, J. B. S. and Balieiro, J. C. C. 2018. Genotype by environment interaction for yearling weight in Nellore cattle applying reaction norms models. Animal Production Science 58:1996-2002. https://doi.org/10.1071/AN17048

Rosa, G. J. M.; Padovani, C. R. and Gianola, D. 2003. Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. Biometrical Journal: Journal of Mathematical Methods in Biosciences 45:573-590. https://doi.org/10.1002/bimj.200390034

Samorè, A. B. and Fontanesi, L. 2016. Genomic selection in pigs: state of the art and perspectives. Italian Journal of Animal Science 15:211-232. https://doi.org/10.1080/1828051X.2016.1172034

Santana, M. L.; Eler, J. P.; Cardoso, F. F.; Albuquerque, L. G. and Ferraz, J. B. S. 2013. Phenotypic plasticity of composite beef cattle performance using reaction norms model with unknown covariate. Animal 7:202-210. https://doi.org/10.1017/S1751731112001711

Sargolzaei, M.; Chesnais, J. P. and Schenkel, F. S. 2011. FImpute-An efficient imputation algorithm for dairy cattle populations. Journal of Dairy Science 94(E-Suppl. 1):421.

Schaeffer, L. R. and Kennedy, B. W. 1986. Computing strategies for solving mixed model equations. Journal of Dairy Science 69:575-579. https://doi.org/10.3168/jds.S0022-0302(86)80441-6

Sorensen, D. and Gianola, D. 2007. Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer, New York, NY.

Strabel, T.; Misztal, I. and Bertrand, J. K. 2001. Approximation of reliabilities for multiple-trait model with maternal effects. Journal of Animal Science 79:833-839. https://doi.org/10.2527/2001.794833x

Strandén, I. and Gianola, D. 1998. Attenuating effects of preferential treatment with Student-t mixed linear models: a simulation study. Genetics Selection Evolution 30:565-583. https://doi.org/10.1186/1297-9686-30-6-565

Strandén, I. and Lidauer, M. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. Journal of Dairy Science 82:2779-2787. https://doi.org/10.3168/jds.S0022-0302(99)75535-9

Tsuruta, S.; Misztal, I. and Stranden, I. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. Journal of Animal Science 79:1166-1172. https://doi.org/10.2527/2001.7951166x

Van Vleck, L. D. and Cassady, J. P. 2004. Modification of MTDFREML to estimate variance due to genetic competition effects. Journal of Animal Science 82(Suppl. 2):38.

Vandenplas, J.; Eding, H.; Calus, M. P. L. and Vuik, C. 2018. Deflated preconditioned conjugate gradient method for solving single-step BLUP models efficiently. Genetics Selection Evolution 50:51. https://doi.org/10.1186/s12711-018-0429-3

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science 91:4414-4423. https://doi.org/10.3168/jds.2007-0980

Weller, J. I.; Ezra, E. and Ron, M. 2017. Invited review: A perspective on the future of genomic selection in dairy cattle. Journal of Dairy Science 100:8633-8644. https://doi.org/10.3168/jds.2017-12879

Westell, R. A.; Quaas, R. L. and Van Vleck, L. D. 1988. Genetic groups in an animal model. Journal of Dairy Science 71:1310-1318. https://doi.org/10.3168/jds.S0022-0302(88)79688-5

Wiggans, G. R.; Cole, J. B.; Hubbard, S. M. and Sonstegard, T. S. 2017. Genomic selection in dairy cattle: the USDA experience. Annual Review of Animal Biosciences 5:309-327. https://doi.org/10.1146/annurev-animal-021815-111422

Zaabza, H. B.; Van Tassell, C. P.; Vandenplas, J.; VanRaden, P. M.; Liu, Z.; Eding, H.; McKay, S.; Haugaard, K.; Lidauer, M. H.; Mäntysaari, E. A. and Strandén, I. 2023. Invited review: Reliability computation from the animal model era to the single-step genomic model era. Journal of Dairy Science 106:1518-1532. https://doi.org/10.3168/jds.2022-22629

Zystro, K.; Peters, T.; Miller, K. and Tracy, W. F. 2021. Classical and genomic prediction of hybrid sweet corn performance in organic environments. Crop Science 61:1698-1708. https://doi.org/10.1002/csc2.20400