sciendo

Contemporary Agriculture, Vol. 72, No. 4 (2023): 251-260
DOI: 10.2478/contagri-2023-0034
UDC: 637.1

Original scientific paper

HERITABILITY AND REPEATABILITY ESTIMATES FOR MILK PRODUCTION TRAITS USING PHENOTYPIC, PEDIGREE-BASED AND GENOMIC DATA OF SERBIAN HOLSTEIN COWS

ISSN 2466-4774
https://www.contagri.info/

OPEN ACCESS

LJUBA ŠTRBAC[1*], *NEBOJŠA DEDOVIĆ*[1], *SNEŽANA TRIVUNOVIĆ*[1], *DOBRILA JANKOVIĆ*[1], *MOMČILO ŠARAN*[1], *DRAGAN STANOJEVIĆ*[2], *RADICA ĐEDOVIĆ*[2], *DONI PRACNER*[3]

[1]University of Novi Sad, Faculty of Agriculture, Trg Dositeja Obradovića 8, 21 000 Novi Sad, Serbia
[2]University of Belgrade, Faculty of Agriculture, Nemanjina 6, 11 080 Zemun-Belgrade, Serbia
[3]University of Novi Sad, Faculty of Science, Trg Dositeja Obradovića 3, 21 000 Novi Sad, Serbia
**Corresponding author*: ljuba.strbac@stocarstvo.edu.rs

Submitted: 09.10.2023.
Accepted: 11.12.2023.

SUMMARY

This research aims to estimate heritability and repeatability based on the data on milk production traits (MY – milk yield; FY – milk fat yield; FC – milk fat content; PY – milk protein yield and PC – milk protein content) as well as pedigree and genomic information. A total of 6,041 animals were included in the research, while 2,565 of them had data for milk production traits. In order to form a genomic relationship matrix, 58K SNP data were used for a total of 1,491 cows. Several software tools were used in the preparation and analysis of data, which were provided by the Central Breeding Organization, Department of Animal Science, Faculty of Agriculture, University of Novi Sad. PreGSF90, in combination with RENUMF90, was used for quality control of genomic information. Genetic analysis was performed in WOMBAT software by the REML using standard repeatability univariate analysis (BLUPpe) and repeatability models for genomic prediction (GBLUPpe and ssGBLUPpe). In all three analyses, the highest heritability (0.410, 0.378 and 0.389, respectively) and repeatability (0.449, 0.429 and 0.440, respectively) were calculated for FC. Heritability estimates for all other traits were lower. Heritability ranged from 0.158 to 0.185 for MY, from 0.166 to 0.178 for FY, from 0.141 to 0.154 for PY and from 0.135 to 0.221 for PC. Heritability estimates indicate that it is possible to achieve genetic improvement but it is necessary to introduce the best model for prediction of breeding values of cow.

Key words:
heritability, milk production traits, pedigree data, SNP data

Abbreviations:
MY – milk yield; FY – milk fat yield; FC – milk fat content; PY – milk protein yield and PC – milk protein content; SNP – single nucleotide polymorphism; REML – restricted maximum likelihood; BLUPpe – Best Linear Unbiesed Prediction with permanent effect; GBLUPpe – genomic BLUPpe; ssGBLUPpe – single step GBLUPpe

INTRODUCTION

With the growth of the world population, it has become necessary to increase the quantity and quality of animal products for human consumption. This is also very important for our country, which has a relatively small number of dairy cattle per capita, while cow's milk and dairy products are among the most common foods in people's diet. Many factors have contributed to the tremendous improvement in dairy cattle production over the past century, while one of the most important factors is regular measuring of traits and recording of phenotypic performance. The quantity and quality of milk are among the key factors affecting profitability of dairy cattle farming and represent the most important characteristics in the breeding programs of various dairy breeds. The most famous and widespread dairy breed in the world is Holstein Friesian. In dairy cattle breeding in Serbia, Holstein Friesian breed of cows is dominant on the territory of AP Vojvodina. According to the data from the Central Breeding Organization for 2022, the number of Holstein Friesian cattle included in breeding programs was 81,554, while the milk yield of the cattle in standard lactation was 7,571 kg of milk, 290 kg of fat (3.83%) and 247 kg of protein (3.26%) (Central Breeding Organization, 2023). By comparing these values with the milk yield of recorded cows in other ICAR member countries (ICAR, 2022), it can be concluded that there are opportunities for improving the traits in the Serbian dairy population. Dairy farmers can improve the productivity of their production through selective breeding. Selective breeding aims to provide offspring that are superior to their parents in traits of interest. Therefore, it is very important that animals are ranked and selected based on their breeding values, which are estimated using various mathematical and statistical models. Development of breeding programs and evaluation of breeding values requires knowledge of genetic parameters, so the first step in carrying out genetic analyzes is their calculation. The most important genetic parameter is heritability, primarily because it predicts how a trait will respond to selection. Dairy cattle breeding programs must consider the repeatable performance of cows, i.e. the potential for more than one lactation per cow (Sahin et al., 2012). The authors state that lifetime milk production is an important economic trait when defining breeding goals and that a very important question is whether milk production in subsequent lactations is sufficiently genetically repeatable, so that production in the first lactation can contribute with useful genetic information on subsequent lactations. The answer to this question leads us to another important genetic parameter, repeatability.

Genetic parameters are calculated by relating certain variance components, and the most popular method for their estimation is the method of restricted maximum likelihood (REML; Patterson & Thompson, 1971). REML theory is based on equations mixed model BLUP (Henderson, 1975), which provides unbiased predictions if all data used for selection are also used in the analysis (Cesarani et al., 2019). When genomic information is available, variance components can be estimated using genome-wide REML (GBLUP) if only the phenotypes of the genotyped animals are considered, or by applying one-step GBLUP (ssGBLUP) if considering the phenotypes of all available animals, including also non-genotyped ones. While the initial cost of REML with genomic data was high due to the dense blocks of mixed equation models generated by the genomic information, updated sparse matrix techniques allow the use of large numbers of genotyped individuals in the estimation of variance components (Masuda et al., 2015). Gutierrez-Reinoso et al. (2021) state that evaluation of various traits in dairy cattle is more efficient using data obtained from genomic analysis compared to pedigree-based analysis, and that combination of both methodologies can significantly improve the evaluation in terms of accuracy.

The main dairy producing countries, including the USA, Canada, Great Britain, Ireland, New Zealand, Australia, France, the Netherlands, Germany and Scandinavian countries, have implemented genomic evaluations in their breeding programs, which has led to significant changes in the global dairy industry sector (Weller et al., 2017). In the breeding program for the Holstein Friesian breed in AP Vojvodina, the genetic evaluation of cattle is carried out on the data on the established lactation production of cows, obtained during the first and later parities using the traditional BLUP sire model and BLUP animal model (Central Breeding Organization, 2019). The aim of this paper is to calculate the heritability and repeatability scores by REML method, using standard univariate analysis with repeated measures (BLUPpe) and repeatability models for genomic prediction (GBLUPpe and ssGBLUPpe – singl step GBLUPpe), and to consider the possibility of introducing genomic information into genetic evaluations.

MATERIAL AND METHODS

Data files

Database consisted of 1,600 Holstein cows genotyped with the GeneSeek GGP Bovine 100 K chip, on the basis of which the preparation of the SNP file started. The first preparation of SNPs involved excluding the chips for so-called Mendelian traits, as well as double SNPs with a smaller GeneTrain parameter. It was followed by the preparation of SNPs for analysis using Wombat software (Meyer, 2007), which does not accept markers with missing values. For this purpose, some SNPs and animals were eliminated, after which the database consisted of

1,566 animals with 62,807 SNPs, each with a call rate of 100%. The next step was SNP quality control in the preGSF90 software in combination with RENUMF90 (Misztal et al., 2018), where one SNP was excluded according to the Mendelian conflict (MC) criterion and 4,091 SNPs with a minor allele frequency <0.05. Also, 14 animals with Mendelian conflicts between parents and offspring were removed. After the quality control, the database included 1,491 cows with 58,715 SNPs.

Based on these cows, three pedigree files with two generations were formed. The first pedigree file contained data for calculating the numerical relationship matrix (A) and was organized in three columns: animal, sire, dam. The second pedigree file was created to calculate a genomic relationship matrix and contained data only for genotyped animals (G). It had the following structure: three columns for parentage information (animal, sire, dam) and a fourth indicating that there is SNP information for that animal. The third pedigree file was a combination of the previous two (H) and it contained data on the origin and genotyping of animals from the A pedigree file with the same data structure as in the G file except that in the fourth column 0 indicated animals that do not have SNPs, and 1 those who do.

After the preparation of SNP and pedigree files, we started to create files with phenotypic data for five milk yield traits (MY – milk yield; FY – milk fat yield; FC – milk fat content; PY – milk protein yield and PC – milk protein content) and factors that were a fixed part of the model for heritability and repeatability assessment. Phenotypic data were related to the production results of lactation standardized to 305 days using the ICAR method to calculate daily values based on monthly controls during AM/PM milking (Delorenzo et al., 1986; ICAR, 2022a). The influence of lactation in order was observed through 6 classes (first lactation - 2565 measurements; second - 1448 measurements; third - 768 measurements; fourth - 344; fifth - 116; sixth - 40). Lactations refer to calvings in the period from 2012 to 2021 and are grouped into three seasons (first - November, December, January, February; second - March, April, September, October; third - May, June, July, August), with a total of 8 farms on the territory of AP Vojvodina.

Table 1 shows an overview of the number of animals included in the input files that participated in the trials.

Table 1. Structure of pedigree files

Number of Animals	Pedigree Files		
	A	***G***	***H***
Total	6041	1491	6041
Full pedigree	4087	1491	4087
Measurement results	2565	1491	2565
Genotyped	-	1491	1491

Legend: *A* – file for the NRM matrix of relationships; *G* and *H* – files for the GRM matrix of relationships

Evaluation of variance components, heritability and repeatability

BLUP was first developed by Henderson (1975) and it is used to simultaneously estimate fixed effects and breeding values. BLUP has found wide use in the genetic evaluation of domestic animals due to its statistical properties. With the constant progress of computers, its application has expanded from simpler models, such as the sire model, to more complex models, such as the animal model and its derivatives. Here, the repeatability models will be presented. In all the presented models, the values suggested by Wombat (Mayer, 2007) were used for the initial values of the variance components.

Let us consider the following equation for a mixed linear model (BLUP):

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e} \quad (1)$$

where are:

$\boldsymbol{y}$ - the vector of observations (features) dimensions n x 1, and n is the number of data
$\boldsymbol{b}$ - the vector of unknown fixed effects of dimension p x 1, and p is the number of levels of fixed effects
$\boldsymbol{a}$ - the vector of unknown random effects (breeding values) of dimension q x 1, and q is the number of levels of random effects
$\boldsymbol{e}$ - the vector of random error effects of dimension n x 1
$\boldsymbol{X}$ - a known design matrix of dimension n x p relating the given data to the fixed effects
$\boldsymbol{Z}$ - a known design matrix of dimension n x q relating the given data to the random effects.

Matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ are also called incidence matrices and their elements are 0 and 1.

The vector a contains only the additive random effects of the animal. Then

$$var(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2,\ var(\boldsymbol{a}) = \boldsymbol{A}\sigma_a^2,\ cov(\boldsymbol{e},\boldsymbol{a}) = 0$$

and

$$var(\boldsymbol{y}) = var(\boldsymbol{Za} + \boldsymbol{e}) = \boldsymbol{Z}var(\boldsymbol{a})\boldsymbol{Z}' + var(\boldsymbol{e}) = \boldsymbol{ZAZ}'\sigma_a^2 + \boldsymbol{I}\sigma_e^2$$

$\boldsymbol{Z}'$ denotes the transposed matrix $\boldsymbol{Z}$. Mathematically speaking, BLUP provides an estimate for $\boldsymbol{b}$ that we denoted by $\widehat{\boldsymbol{b}}$ and an estimate for $\boldsymbol{a}$, denoted by $\widehat{\boldsymbol{a}}$, which were obtained by solving the equations (MME - Mixed Model Equations) which are, for Equation (1), given as

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{\alpha_1}\boldsymbol{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{y} \end{bmatrix} \quad (2)$$

where $\alpha_1 = \sigma_e^2/\sigma_\alpha^2$.

The starting values of the variance components (Henderson-s method 3)
The Least Squares equations for (1) are (van der Werf, 2009)

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{y} \end{bmatrix} \quad (3)$$

Absorbing the fixed effects reduces the equations to $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}\widehat{\boldsymbol{a}} = \boldsymbol{Z}'\boldsymbol{M}\boldsymbol{y}$ with $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. If the inverse of $\boldsymbol{X}'\boldsymbol{X}$ does not exist, a generalized inverse can be used in its place. The initial values of the variance components can be taken as

$$\sigma_e^2 = \frac{\boldsymbol{y}'\boldsymbol{y} - \widehat{\boldsymbol{a}}'\boldsymbol{Z}'\boldsymbol{y} - \widehat{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{y}}{n - rank(\boldsymbol{X}) - rank(\boldsymbol{Z}) + 1}$$

and

$$\sigma_a^2 = \frac{\widehat{\boldsymbol{a}}'\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{y} - (rank(\boldsymbol{Z}) - 1)\sigma_e^2}{trace(\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})}$$

The Average Information Algorithm for Restricted Maximum Likelihood (AI-REML) estimation
Basic steps of this AI-REML are:
Step 1: Determine the NRM matrix of relationships marked as $\boldsymbol{A}$.
Step 2: Set the initial values of the variances σ_a^2 and σ_e^2, set the counter $k = 0$ and define $\boldsymbol{\theta}_0 = \begin{bmatrix} \sigma_e^2 \\ \sigma_a^2 \end{bmatrix}$.
Step 3: Calculate $\alpha_1 = \sigma_e^2/\sigma_a^2$.
Step 4: Determine the matrices $\boldsymbol{R} = \sigma_e^2\boldsymbol{I}$ and $\boldsymbol{G} = \sigma_a^2\boldsymbol{A}$.
Step 5: Define the matrices $\boldsymbol{V} = \boldsymbol{ZGZ}' + \boldsymbol{R}$ and $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}$.
Step 6: If we denote with $|*|$ determinant of the matrix, we calculate $\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}$, $log|\boldsymbol{V}|$ and $log|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}|$.
Step 7: Calculate the value of the likelihood function

$$L(\sigma_e^2, \sigma_a^2) = \frac{1}{2}(-\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y} - log|\boldsymbol{V}| - log|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}|)$$

Step 8: Solve the system (2).
Step 9: With $\boldsymbol{C}^{22}$ we denote the part of the inverse matrix from Step 8, which corresponds to unknown random effects. Let $\boldsymbol{s} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}} - \boldsymbol{Z}\widehat{\boldsymbol{a}}$, where $\widehat{\boldsymbol{b}}$ and $\widehat{\boldsymbol{a}}$ are obtained from Step 8. Calculate partial derivative of the first order as follows:

$$\frac{\partial L}{\partial \sigma_e^2} = \frac{1}{2}\left(\frac{\boldsymbol{s}'\boldsymbol{s}}{(\sigma_e^2)^2} - \frac{n - p - q}{\sigma_e^2} - \frac{trace(\boldsymbol{C}^{22}\boldsymbol{A}^{-1})}{\sigma_a^2}\right)$$

$$\frac{\partial L}{\partial \sigma_a^2} = \frac{1}{2}\left(\frac{\hat{\boldsymbol{a}}' \boldsymbol{A^{-1}} \hat{\boldsymbol{a}}}{(\sigma_a^2)^2} - \frac{q}{\sigma_a^2} + \frac{trace(\boldsymbol{C^{22}A^{-1}})\sigma_e^2}{(\sigma_a^2)^2}\right)$$

Step 10: Calculate the average (*av*) partial derivative of the second order as follows:

$$M = av\left(\frac{\partial^2 L}{(\partial \sigma_e^2)^2}\right) = -\frac{1}{2}\left(\frac{\boldsymbol{s'Ps}}{(\sigma_e^2)^2}\right)$$

$$N = av\left(\frac{\partial^2 L}{(\partial \sigma_a^2)^2}\right) = -\frac{1}{2}\left(\frac{\hat{\boldsymbol{a}}' \boldsymbol{Z'PZ} \hat{\boldsymbol{a}}}{(\sigma_a^2)^2}\right)$$

$$Q = av\left(\frac{\partial^2 L}{\partial \sigma_e^2 \partial \sigma_a^2}\right) = -\frac{1}{2}\left(\frac{\hat{\boldsymbol{a}}' \boldsymbol{Z'Ps}}{\sigma_e^2 \sigma_a^2}\right)$$

Step 11: Form the matrix $\boldsymbol{A}_{inf} = \begin{bmatrix} -M & -Q \\ -Q & -N \end{bmatrix}$ and calculate $\boldsymbol{A}_{inf}^{-1}$.

Step 12: Now, calculate new estimations of the variance components:

$$\begin{bmatrix} \sigma_e^{2\,2} \\ \sigma_{\alpha\,e}^2 \end{bmatrix} = \boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \boldsymbol{A}_{inf}^{-1} \cdot \begin{bmatrix} \partial \mathrm{L} / \partial \sigma_{\mathrm{e}}^2 \\ \partial \mathrm{L} / \partial \sigma_{\alpha}^2 \end{bmatrix}$$

Step 13: If the difference between variance components calculated in two successive iterations is small enough, we calculate the standard errors (*SE*) by formulas:

$$SE_{\sigma_e^2} = \sqrt{\boldsymbol{A}_{inf}^{-1}(1,1)},\ SE_{\sigma_a^2} = \sqrt{\boldsymbol{A}_{inf}^{-1}(2,2)}$$

Otherwise, $k = k + 1$ and go to Step 3.

The similar algorithm can be applied to a mixed linear model with repetition. Let us consider the following equation for a mixed linear model with repetition (BLUPpe):

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{Wpe} + \boldsymbol{e} \tag{4}$$

where are:

pe - the vector of unknown environmental random effects of dimension *r* x 1; the matrix $\boldsymbol{W}$ is a design matrix related to the effects of the external environment (i.e., only the animals for which we had repeated measurements were observed) of dimension *n* x *r*.

The vector a contains only the additive random effects of the animal, while the non-additive genetic effects are contained in the vector ***pe***. The assumption is that environmental effects and error effects are independently distributed with a mean of zero and variances σ_{pe}^2 and σ_e^2, respectively. That is when

$$var(\boldsymbol{pe}) = \boldsymbol{I}\sigma_{pe}^2,\ var(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2,\ var(\boldsymbol{a}) = \boldsymbol{A}\sigma_a^2$$

and

$$var(\boldsymbol{y}) = var(\boldsymbol{Za} + \boldsymbol{Wpe} + \boldsymbol{e}) = \boldsymbol{Z}var(\boldsymbol{a})\boldsymbol{Z}' + \boldsymbol{W}var(\boldsymbol{pe})\boldsymbol{W}' + var(\boldsymbol{e})$$

$$= \boldsymbol{ZAZ}'\sigma_a^2 + \boldsymbol{WW}'\sigma_{pe}^2 + \boldsymbol{I}\sigma_e^2$$

$\boldsymbol{W}'$ denotes the transposed matrix $\boldsymbol{W}$. Mathematically speaking, BLUPpe provides an estimate for $\boldsymbol{b}$ that we denoted by $\hat{\boldsymbol{b}}$, an estimate for $\boldsymbol{a}$, denoted by $\hat{\boldsymbol{a}}$, and an estimate for ***pe***, denoted by $\widehat{\boldsymbol{pe}}$ and which were obtained by solving the equations (MME - Mixed Model Equations) which are, for Equation (4), given as

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} & \boldsymbol{X'W} \\ \boldsymbol{Z'X} & \boldsymbol{Z'Z + \alpha_1 A^{-1}} & \boldsymbol{Z'W} \\ \boldsymbol{W'X} & \boldsymbol{W'Z} & \boldsymbol{W'W + \alpha_2 I} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{b}} \\ \widehat{\boldsymbol{a}} \\ \widehat{\boldsymbol{pe}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{Z'y} \\ \boldsymbol{W'y} \end{bmatrix} \quad (5)$$

where $\alpha_1 = \frac{\sigma_e^2}{\sigma_a^2}$ and $\alpha_2 = \frac{\sigma_e^2}{\sigma_{pe}^2}$.

GBLUP is BLUP, where $\boldsymbol{A}$ is replaced by the genomic matrix $\boldsymbol{G}$, i.e., $\boldsymbol{G_r}$ (for more details about matrices $\boldsymbol{G}$, i.e. $\boldsymbol{G_r}$, see Štrbac et al., 2023). The basic model was then Equation (4) with $var(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2$, $var(\boldsymbol{a}) = \boldsymbol{G}\sigma_a^2$ and $var(\boldsymbol{pe}) = \boldsymbol{I}\sigma_{pe}^2$. The MME model with repetition (GBLUPpe) is

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} & \boldsymbol{X'W} \\ \boldsymbol{Z'X} & \boldsymbol{Z'Z + \alpha_1 G^{-1}} & \boldsymbol{Z'W} \\ \boldsymbol{W'X} & \boldsymbol{W'Z} & \boldsymbol{W'W + \alpha_2 I} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{b}} \\ \widehat{\boldsymbol{a}} \\ \widehat{\boldsymbol{pe}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{Z'y} \\ \boldsymbol{W'y} \end{bmatrix} \quad (6)$$

where $\alpha_1 = \frac{\sigma_e^2}{\sigma_a^2}$ and $\alpha_2 = \frac{\sigma_e^2}{\sigma_{pe}^2}$. If the matrix in Equations (6) is singular, its generalized inverse matrix is used when obtaining $\widehat{\boldsymbol{b}}$, $\widehat{\boldsymbol{a}}$ and $\widehat{\boldsymbol{pe}}$.

The idea for ssGBLUP arose from the fact that a smaller part of the animals from the observed population was genotyped. A relationship matrix of the form $\boldsymbol{A}$ is observed

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{bmatrix}$$

where index number 1 refers to animals that have not been genotyped, and index number 2 to animals that have been genotyped. Now, the matrix $\boldsymbol{M_{g012}}$ has dimensions q_g x m, where q_g is the number of genotyped animals and m is the number of marker columns (i.e., the number of SNPs). After centering the matrix $\boldsymbol{M_{g012}}$, we get the matrix $\boldsymbol{M_g}$, whose elements are −1, 0 and 1. After determining the allele frequency (as in the previous section), the matrix $\boldsymbol{S}$ will be of the form $\boldsymbol{S} = M_g - \boldsymbol{P}$. The genomic relationship matrix ($\boldsymbol{G_g}$) is given by

$$\boldsymbol{G_g} = \frac{\boldsymbol{SS'}}{2\sum_{j=1}^{m_g} p_{j,B}(1 - p_{j,B})}$$

where are:

$p_{j,B}$ - the frequency of allele B in the j-th column of the SNP. If the matrix $\boldsymbol{G_g}$ is singular, then it is replaced by $\boldsymbol{G_{gr}}$ where

$$\boldsymbol{G_{gr}} = \lambda\left(\beta\left(\boldsymbol{G_g} + \varepsilon\boldsymbol{I}\right) + \alpha\boldsymbol{J}\right) + (1 - \lambda)\boldsymbol{A_{22}}$$

Here, with $0 \leq \lambda \leq 1$ we denoted the ratio of the total genetic variance and the marker effect, $\boldsymbol{A}$ as the relationship matrix, α, and β as the leveling factors proposed in [23] and [24], $\boldsymbol{J}$ is a square matrix with all elements equal to 1, $\boldsymbol{I}$ is a unit matrix and $0 < \varepsilon \ll 1$ is a small constant. Clearly, $\boldsymbol{A_{22}}$ is a matrix of dimension q_g x q_g and is the relationship matrix only for genotyped animals. Matrices $\boldsymbol{J}$ and $\boldsymbol{I}$ were defined in the previous section. The matrix $\boldsymbol{H^{-1}}$is obtained as follows [25]

$$\boldsymbol{H^{-1}} = \boldsymbol{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G_{gr}^{-1} - A_{22}^{-1}} \end{bmatrix}.$$

For ssGBLUPpe, the basic model is Equation (4), with $var(\boldsymbol{e}) = \boldsymbol{I}\sigma_e^2$, $var(a) = \boldsymbol{H}\sigma_a^2$ and $var(\boldsymbol{pe}) = \boldsymbol{I}\sigma_{pe}^2$. The MME model for Equation (4) with repetition is

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'Z} & \boldsymbol{X'W} \\ \boldsymbol{Z'X} & \boldsymbol{Z'Z + \alpha_1 H^{-1}} & \boldsymbol{Z'W} \\ \boldsymbol{W'X} & \boldsymbol{W'Z} & \boldsymbol{W'W + \alpha_2 I} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{b}} \\ \widehat{\boldsymbol{a}} \\ \widehat{\boldsymbol{pe}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{Z'y} \\ \boldsymbol{W'y} \end{bmatrix} \quad (7)$$

where $\alpha_1 = \frac{\sigma_e^2}{\sigma_a^2}$ and $\alpha_2 = \frac{\sigma_e^2}{\sigma_{pe}^2}$. In the case of a singular matrix for this model, its generalized inverse matrix was used. Note that if all animals are genotyped, $\boldsymbol{H^{-1} = G_{gr}^{-1}}$.

The phenotypic variance $\sigma_y^2 = \sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2$, while the repeatability coefficient is equal to $R = (\sigma_a^2 + \sigma_{pe}^2)/\sigma_y^2$. The heritability is $h^2 = \sigma_a^2/\sigma_y^2$.

RESULTS AND DISCUSSION

In most of REML algorithms, the iterations are used. The algorithm starts by the starting values of the variance components and finishes when the likelihood function reaches its maximum. The basic properties of this algorithm are: partial derivatives of the maximum likelihood function are needed; the average of partial derivatives of the maximum likelihood function of the second order and the expected values of these partial derivatives must be calculated; it includes unknown random effects; at each iteration, the solution of MME and trace of part of the inverse matrix which corresponds to unknown random effects; and algorithm stops when the difference between the calculated variance components is less than some given number. In all the presented models that have been considered, the values suggested by Wombat (Meyer, 2007) were used for the starting values of the variance components (Tab. 2).

Table 2. The starting values of the variance components when solving system (2), (3) and (4)

Method		**MY**	**FY**	**FC**	**PY**	**PC**
BLUPpe	$\sigma_e^2 = \sigma_a^2 = \sigma_{pe}^2$	3939814	8442	0.4442	3990	0.0568
GBLUPpe		3656890	8223	0.4620	3981	0.0662
ssGBLUPpe		3939814	8442	0.4442	3990	0.0568

From the data in Table 2, we can see that the same starting values were recommended by the software during BLUPpe and ssGBLUPpe, which is explained by the fact that the number of animals with measurement results was the same, while in GBLUPpe the phenotypic values of only genotyped animals were used, i.e., there were fewer values.

The goal of animal breeding is to improve animal performance in specific populations through selection. This improvement would not be possible without variability, i.e., differences between animals that are quantified at the population level by estimating the variance components for the traits of interest. Variance components are important to us for evaluation of genetic parameters and evaluation of breeding values. Because of its desirable traits, REML has become the most widely applied method for estimating variance components using different BLUP models. Also, the availability of modern high-capacity computers and the progress in increasing the efficiency of computer algorithms make it possible to test more and more complex models for this purpose, and therefore the development and application of appropriate statistical procedures for comparing models is becoming increasingly important (Hofer, 1998). The author points out the algorithms that are based on (approximate) second-order partial derivatives, such as average information (AI) REML or quasi-Newton, provide approximate large sample (co)variances of estimated (co)variance components.

Table 3 shows the estimates of variance components and standard errors of estimates calculated by the REML method within the traditional BLUP model with repeated measures and BLUP models with genomic data.

Table 3. Evaluation of variance components

		$\boldsymbol{\sigma_{a-SNP}^2 \pm s.e.}$	$\boldsymbol{\sigma_{pe}^2 \pm s.e.}$	$\boldsymbol{\sigma_e^2 \pm s.e.}$
BLUPpe	MY	448040 ± 95779.1	652881 ± 89324.9	1740830 ± 47068.5
	FY	1006 ± 186.5	1014 ± 171.5	3639 ± 98.4
	FC	0.102 ± 0.0103	0.010 ± 0.0074	0.138 ± 0.0038
	PY	400.5 ± 89.80	592.4 ± 85.21	1783 ± 48.1
	PC	0.011 ± 0.0017	0.006 ± 0.0015	0.034 ± 0.0009

GBLUPpe	MY	480433 ± 112529.0	446426 ± 113878.0	1951760 ± 80670.8
	FY	988 ± 220.72	869.1 ± 230.5	4103 ± 171.0
	FC	0.093 ± 0.0126	0.013 ± 0.0010	0.141 ± 0.006
	PY	415.2 ± 108.68	423.1 ± 115.6	2108 ± 86.4
	PC	0.008 ± 0.0020	0.005 ± 0.0022	0.048 ± 0.0019
ssGBLUPpe	MY	568693 ± 91689.6	492457 ± 83453.4	2007740 ± 54001.7
	FY	1050 ± 167.0	887.7 ± 156.99	3982 ± 107.0
	FC	0.096 ± 0.0091	0.013 ± 0.0065	0.138 ± 0.0039
	PY	463.6 ± 83.28	460.4 ± 80.14	2079 ± 55.8
	PC	0.011 ± 0.0014	0.006 ± 0.0013	0.034 ± 0.0009

Variance component estimates were in most cases higher when genomic data were included, except in the case of variance estimates due to environmental permanent effects, which were highest using the traditional BLUPpe model. From the example of σ^2_{a-SNP} for the trait MY, we can see that the calculated scores increased by about 7% and 27% when we compare the values obtained by applying GBLUP and ssGBLUP in relation to BLUPpe, respectively. Regarding the milk quality traits, different results were obtained. GBLUP gave a higher σ^2_{a-SNP} for the trait PY, while for the other three traits the values were lower compared to BLUPpe. Interestingly, ssGBLUPpe gave the opposite results, i.e., higher values of σ^2_{a-SNP} were for all milk quality traits except for the PY trait. By comparing the model with the genomic data, we can see that the values were higher for all five analyzed traits in favor of ssGBLUP. As for σ^2_e, using the model with genomic data, higher scores were obtained for almost all traits, except for the FC and PC traits, where the same values of this variance were obtained using ssGBLUPpe and BLUPpe.

If we compare the standard error values of the estimates of the variance components, we can see that higher values were obtained using GBLUPpe compared to the traditional BLUPpe, except for the FC trait, and by comparing the values obtained from ssGBLUPpe with BLUPpe, we can see that the standard errors were smaller, except for σ^2_e. Standard errors obtained by models with genomic data were smaller in favor of ssGBLUPpe.

Heritability is the most important genetic parameter, on the basis of which breeding programs define methods for evaluating the breeding values of animals, selection methods and mating systems. In addition to heritability, which can be interpreted as a parameter that assesses to what extent the superiority of the parents can be expected in their offspring, it is also very important to assess repeatability, which assesses to what extent the superiority of animals in one measurement can be expected in subsequent measurements of the same animals. Estimates of heritability and repeatability in populations depend on partitioning the observed variation into components that reflect genetic and environmental factors. Although these data are most commonly calculated using traditional methods based on pedigree and phenotype data, new methods using genetic marker data have been recently proposed. Table 4 shows heritability and repeatability scores with standard errors of heritability scores.

Table 4. Evaluation of heritability (h^2) and repeatability (R)

Method	**BLUPpe**		**GBLUPpe**		**ssGBLUPpe**	
Traits	h^2**±*s.e.***	***R***	h^2**±*s.e.***	***R***	h^2**±*s.e.***	***R***
MY	0.158 ± 0.032	0.387	0.167 ± 0.037	0.322	0.185 ± 0.028	0.346
FY	0.178 ± 0.031	0.357	0.166 ± 0.035	0.312	0.177 ± 0.027	0.327
FC	**0.410** ± 0.035	0.449	**0.378** ± 0.044	0.429	**0.389** ± 0.031	0.440
PY	0.144 ± 0.031	0.358	0.141± 0.036	0.285	0.154 ± 0.026	0.308
PC	0.221 ± 0.031	0.342	0.135 ± 0.031	0.223	0.216 ± 0.026	0.333

Based on the data from Table 4, we can see that higher heritability values were calculated for the MY trait when genomic data were used, and the highest value was obtained using the ssGBLUP model. The highest value for the trait PY was also calculated using the ssGBLUP model (0.154), but it was slightly higher compared to BLUPpe (0.144) and GBLUPpe (0.141), which gave similar results. For other traits (FY, FC and PC), the highest heritability value was calculated using the traditional BLUPpe model, which also calculated the highest repeatability coefficient values for all observed traits. In all three analyses, the highest heritability scores, which were moderately high (0.410, 0.378 and 0.389, respectively), as well as repeatability scores (0.449, 0.429 and 0.440, respectively) were calculated for the FC trait. For other traits, heritability values were moderate. If we look at the values of standard errors of heritability estimates, we can see that the lowest values were calculated using the ssGBLUPpe model. A little higher values were calculated using the traditional BLUPpe model, while the highest values were calculated using the GBLUPpe model.

If we compare our results with the results in the literature, we can see that there are certain agreements. Sermyagin et al. (2018) performed a genome-wide association study for milk production traits in a Russian population of Holstein and Black and White cattle and determined that the heritability of milk yield was 0.180. The highest heritability was calculated for milk fat yield at 0.221, while for milk protein yield it was 0.173. Based on our research, we can see that a similar heritability value was obtained for the trait MY, which was of medium degree. This may be due to the fact that MY is the trait with the highest selection pressure over a longer period of time compared to other traits where higher heritability values were calculated. In the research by Lee et al. (2020), the estimated heritability of milk yield per parity in the first, second and third parity was 0.28, 0.20 and 0.16, respectively, while for fat yield it was 0.26, 0.23 and 0.20, and for protein yield it was 0.23, 0.18, and 0.15, respectively. The highest heritability estimates for milk fat percentage and protein percentage, as was also the case in our study, were calculated by Oliveira Junior et al. (2021), but their values were much higher (0.66 and 0.69, respectively) and the applied model was a bivariate linear animal model using Bayesian methods via Gibbs sampling.

It is interesting that the heritability value calculated using the traditional BLUPpe model was higher compared to those with genomic data for most traits. However, the case that most of the identified SNPs explain a small part of heritability for certain traits can be found in the literature and other studies, and the difference between heritability based on pedigree and that calculated on genomic data was called missing heritability (Zhu & Zhou, 2020). In the research of Khanzadeh et al. (2022), estimated heritabilities based on pedigree and genome were 0.253 and 0.144 for milk yield, 0.290 and 0.191 for milk fat percentage, and 0.378 and 0.363 for milk protein percentage, respectively. Zhu & Zhou (2020) state that missing heritability may be due to current GWAS being not powerful enough and many so-called causative SNPs remain undetected, which can greatly underestimate the variance estimate. In addition, pedigree-based studies may overestimate heritability, which depends on the applied model, factors and the amount of information included in the analysis.

CONCLUSION

Heritability in the narrow sense is an important genetic parameter that quantifies the proportion of the phenotypic variance of a particular trait that can be attributed to additive genetic variation. Repeatability represents its upper limit because, in addition to genetic variations, it also takes into account variations due to permanent effects of the external environment, i.e., it explains the extent to which phenotypic differences between animals can be explained by permanent variations. Estimation of these parameters previously relied on traditional methods that included pedigree data and phenotopic measurements. The use of genetic markers enabled the development of new methods for evaluating the so-called genomic heritability, which is the proportion of phenotypic variance explained by SNPs.

In this study, variance component estimates were in most cases higher when genomic data were included, except in the case of variance estimates due to environmental permanent effects, which were the highest using the traditional BLUPpe model. Higher heritability values for the trait MY were calculated when genomic data were used, and the highest value was obtained using the ssGBLUP model. The highest heritability was calculated using the ssGBLUP model also for the PY trait, while for the FY, FC and PC traits, the highest heritability values were calculated using the traditional BLUPpe model. The highest values of the repeatability coefficient for all observed traits were calculated using the traditional BLUPpe model. In all three analyses, the trait FC had the highest heritability scores, which were medium high (0.410, 0.378 and 0.389, respectively) as well as repeatability scores (0.449, 0.429 and 0.440, respectively), while the heritability values for the other traits were medium.

Heritability estimates indicate that it is possible to achieve genetic improvement, but it is necessary to introduce the best model for predicting the breeding values of cows. The research presented in this paper indicates that the standard errors of heritability estimates, and in most cases the component variances, were the lowest when applying the ssGBLUPpe model, while the GBLUPpe gave the highest errors. This can be explained by the small number of genotyped animals, namely cows without the genomic information of sires. The main limiting factor for wide application of genomic methods in animal breeding is the insufficient size of the population in certain countries, including ours, so it is necessary to consider combining reference populations in different countries, exchanging genotypes and thereby contributing to improvement of genetic assessments.

Acknowledgements**:** This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No.451-03-47/2023-01/ 200117). The authors express their gratitude to the Cattle Breeding Unit of the Animal Breeding Center within the Department of Animal Science of the Faculty of Agriculture, University of Novi Sad, which is the Central Breeding Organization for the Autonomous Province of Vojvodina, and their staff who kindly provided us with the pedigree and performance data from the Herd Book of Registered Animals. The data for this research were collected during realization of the project BioITGenoSelect (Project number: 6066512) funded by The Science Fund of the Republic of Serbia, through the PROMIS project program.

Conflict of interest: The authors declare that they have no conflict of interest.

REFERENCES

Central Breeding Organization (2022): The Report on the Implementation of Breeding Programs in Livestock of AP Vojvodina for the Year 2021. Available at: https://www.stocarstvo.edu.rs/centar (accessed on 10 September 2023).

Central Breeding Organization (2019): Breeding program for Holstein Frisian breed of cattle in AP Vojvodina. Available at: https://www.og.stocarstvo.edu.rs/op (accessed on 10 September 2023)

Cesarani A., Pocrnic I., Macciotta N.P.P., Fragomeni B.O., Misztal I., Lourenco D.A.L. (2019): Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. *Journal of Animal Breeding and Genetics*, 136: 40-50.

Delorenzo, M.A. & Wiggans, G.R. (1986): Factors for Estimating Dairy Yield of Milk, Fat, and Protein from a Single Milking for Herds Milked Twice a Day. *Journal of Dairy Science*, 69(9): 2386-2394.

Gutierrez-Reinoso M.A., Aponte P.M., Garcia-Herreros M. (2021): Genomic Analysis, Progress and Future Perspectives in Dairy Cattle Selection: A Review. *Animals (Basel)*: 11(3): 599.

Henderson C.R. (1975): Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31: 423-447.

Hofer A. (1998): Variance component estimation in animal breeding: a review, *Animal Breeding and Genetics*, 115(1-6): 247-265.

ICAR (2022): The Global Standard for Livestock Data. Statistics 2022. Available at: https://my.icar.org/stats/list (accessed on 10 September 2023)

ICAR (2022a): The Global Standard for Livestock Data. Section 2 - Guidelines for Diary Cattle Milk Recording. Available at: https://www.icar.org/Guidelines/02-Overview-Cattle-Milk-Recording.pdf (accessed on 10 September 2023).

Khanzadeh H., Ghavi Hossein-Zadeh N., Ghovvati S. (2022): A meta-analysis of the gap between pedigree-based and genomic heritability estimates for production traits in dairy cows. *Livestock Science*, 263: 105000.

Lee Y.M., Dang C.G., Alam M.Z., Kim Y.S., Cho K.H., Park K.D., Kim J.J. (2020): The effectiveness of genomic selection for milk production traits of Holstein dairy cattle. *Asian-Australasian Journal of Animal Sciences*, 33(3): 382-389.

Meyer K. (2007): WOMBAT - A tool for mixed model analyses in quantitative genetics by REML. *Journal of Zhejiang University-Science B*, 8: 815-821.

Misztal I., Tsuruta S., Lourenco D.A.L., Masuda Y., Aguilar I., Legarra A., Vitezica Z. (2018): Manual for BLUPF90 Family Programs, University of Georgia. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation (accessed on 17 June 2023).

Oliveira Junior G.A., Schenkel F.S., Alcantara L., Houlahan K., Lynch C., Baes C.F. (2021): Estimated genetic parameters for all genetically evaluated traits in Canadian Holsteins, *Journal of Dairy Science*, 104(8): 9002-9015.

Patterson H.D. & Thompson R. (1971): Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, 58(3): 545-554.

Sahin A., Ulutas Z., Adkinson A.Y., Adkinson R.W. (2012): Genetic and environmental parameters and trends for milk production of Holstein cattle in Turkey. *Italian Journal of Animal Science*, 11(e4): 242-248.

Sermyagin A.A., Gladyr E.A., Plemyashov K.V., Kudinov A.A., Dotsev A.V., Deniskova T.E., Zinovieva N.A. (2016): Genome-Wide Association Studies for Milk Production Traits in Russian Population of Holstein and Black-and-White Cattle. In: Anisimov, K., et al. Proceedings of the Scientific-Practical Conference "Research and Development - 2016". Springer, Cham.

Štrbac Lj., Pracner D., Šaran M., Janković D., Trivunović S., Ivković M., Tarjan L., Dedović N. (2023): Mathematical Modeling and Software Tools for Breeding Value Estimation Based on Phenotypic, Pedigree and Genomic Information of Holstein Friesian Cattle in Serbia. *Animals*, 13: 597.

Van der Werf J (2009): Estimation of Genetic Parameters. Available at: https://www.woolwise.com/wp-content/uploads/2017/07/GENE-422-522-11-T-07.pdf (accessed 20 August 2023).

Weller J.I., Ezra E., Ron M. (2017): Invited review: A perspective on the future of genomic selection in dairy cattle. *Journal of Dairy Science*, 100: 8633-8644.

Zhu H. & Zhou X. (2020): Statistical methods for SNP heritability estimation and partition: A review. *Computational and Structural Biotechnology Journal*, 18: 1557-1568.