

Possibilities of Using “BUGA” Named Software for Breeding Value Estimation of Anatolian Water Buffalo Population of Istanbul

Kemal YAZGAN^{1,*} , Mehmet İhsan SOYSAL² 

¹ Harran University, Faculty of Agriculture, Department of Animal Science, Sanliurfa, Turkiye.

² Namik Kemal University, Faculty of Agriculture, Department of Animal Science, Tekirdag, Turkiye.

*Corresponding Author

Article History

Received: 22.11.2023

Accepted: 26.12.2023

First Online:

Corresponding Author

E-mail: kyazgan@harran.edu.tr
kemalyazgan@gmail.com

Keywords

(co)variance components
average information algorithm
expectation-maximization algorithm
computer software

Abstract

In this study, variance components and breeding value estimations of Anatolian water buffaloes were estimated from the data obtained within the scope of community based water buffalo improvement program of Anatolian water buffalo population of İstanbul. The "BUGA" named software, which was started to be developed with the support of Harran University Scientific Research Project Unit (HUBAP) for use in the estimation of (co)variance components and breeding value were used in this study. The BUGA software used in the research can predict genetic and environmental effects together with variance components as Best Linear Unbiased Prediction (BLUP) with the Restricted Maximum Log likelihood (REML) method. Unlike its counterparts of another software of foreign origin, it has a visual interface. The language of use is Turkish. The dataset includes records such as city, district, village, genealogy information, date of birth, gender, calving date, lactation parity, number of test day milk yield, lactation milk yield and previous calving date of 442 Anatolian buffaloes raised in this population. The records were analysed using the repeatability model with BUGA software and the additive genetic and permanent environmental effects of all animals in the dataset were estimated. Random environment variance, genotypic variance and permanent environment variance values were estimated as 107112.60 kg², 20045.22 kg² and 1259.50 kg², respectively, while heritability (h²) and degree of repeatability (r) were determined as 0.1561 and 0.1659, respectively. The results obtained as an example of case study applied for the possibility of usage the BUGA software in the estimation of genetic parameters and breeding values showed that BUGA software can be easily used in farm recording conditions with the advantage of friendly user features.

Introduction

The identification and utilization of genetically superior individuals in animal breeding depend on the reliable determination of various genetic parameters, such as variance components, heritability, and breeding value, given the genotypic perspective (Kumlu, 2003). Since this is a highly complex process, more than twenty specialized computer software programs have been developed for this purpose since the 1960s to facilitate fast and easy calculations in this field. Some of these software programs have become obsolete as they are no longer being developed. The most commonly used software programs today include MTDFREML (Boldmon *et al.*, 1995), ASREML (Gilmour *et al.*, 1998), BLUPF90 (Anonymous, 2023a), WOMBAT (Meyer, 2007), and MIXBLUP (Anonymous, 2023b). A common feature of these software programs is that they are written in Fortran 77/90/95, C, and Pascal languages and operate through a command system.

There are a number of steps involved in the selection of animals to maximize the response achieved. These include: (i) managing the animals as equally as possible to make it easier to disentangle genetics and environment; (ii) adjusting records of performance for known environmental effects, and then (iii) predicting the breeding values of individual animals by the most appropriate method. Modern (BLUP) methods of genetic evaluation achieve the second and third of these steps simultaneously (Simm *et al.*, 2022).

Predicting an animal's breeding value is a bit like completing a large, complicated jigsaw puzzle, where each piece of the puzzle is a record of performance from the animal itself or one of its relatives. More recently, the genomic information on the animal adds a further piece of the puzzle. Generally, the higher the proportion of genes in common between the animal and a given relative, the more useful the record of performance from that relative. But, records from progeny are of most value. As the number of records on progeny

increases, the correlation between predicted and true breeding values approaches one. With other classes of relatives the accuracy of prediction never reaches one, and for all classes of relatives there are diminishing returns in accuracy as the number of records increases (Simm *et al.*, 2022).

Estimated breeding values (EBVs) can have positive or negative values or be equal to zero. The sign indicates whether they are expected to be genetically above (+) or below (-) the average of the group of animals on which the calculations were performed, or some other defined group of animals whose predicted breeding values (PBVs) are set to average zero (the base). PBVs are expressed (at least initially) in the same units as the record of performance (e.g. kg of live weight, litres of milk, mm of fat). The PBVs of animals can only be compared within contemporary groups, herds or flocks unless there are genetic links between these groups, and the PBVs were from across-herd or across-flock BLUP evaluations (Mrode, 2005; Simm *et al.*, 2022).

In the simplest case, when we have a single record of performance on the animal itself, the predicted or estimated breeding value (PBV or EBV) is the deviation in performance from contemporaries, multiplied by the heritability of the trait concerned. The deviation in performance is calculated after adjusting the performance records for environmental effects. PBVs calculated from a single record of performance span a narrower range than the deviations in performance; the higher the heritability, the lower the proportion of non-genetic variation, and so the less severe the shrinking.

To run them, a recoded pedigree file and complex parameter files must be prepared. These files should specify which model will be used, which fixed and random effects will be included in the model, the initial values of the variance components to be estimated, and the number of iterations, among other parameters, all coded with great care. Almost all of these programs cannot directly read data organized in MS Excel. Data prepared in MS Excel must be converted to a different format called a "data file." Some of these programs are commercial (ASREML and MIXBLUP), while others are free (MTDFREML, WOMBAT, and BLUPF90) or partially free. These software programs lack user-friendly interfaces, require advanced knowledge for operation, suffer from a lack of Turkish resources explaining how to use them, and, in some cases, are paid. As a result, they are not effectively utilized in the livestock sector in our country.

To address these issues, there is a need for the development of a program that is user-friendly, does not require expertise or complex parameter file preparation, and has a Turkish user guide and interface specifically designed for the livestock sector and researchers. As a result of a project supported by Harran University Scientific Research Project Unit (HUBAP) (project no: 22035), a software program named BUGA has been developed. The aim of this study is to demonstrate the usability of the BUGA software in field conditions.

Material and Methods

Data

The data used in the research was obtained from a private enterprise affiliated with the Breeding Buffalo Farmers Union in Istanbul. The dataset includes records for 442 healthy Anatolian buffaloes, such as the city, district, village, pedigree information, date of birth, gender, calving date, lactation order, number of test days, lactation milk yield, and previous calving date.

Statistical Model:

For the determination of variance components, the following model (Equation 1) was preferred for the repeated observations used in the study (Mrode, 2005). The effects of environmental factors, such as birth year and month, lactation order, calving year and month, and lactation duration, were determined using variance analysis through the SAS software (2000), with the results showing that the effects of lactation order, calving month, and lactation duration were significant ($P < 0.05$), while the effects of other factors were deemed non-significant ($P > 0.05$). The significant environmental factors were used as fixed-effect environmental factors in the animal model containing repeated observations.

$$y = Xb + Za + Wpe + e \quad [1]$$

where:

y : vector of observations,

b : vector of fixed effects,

a : vector of random animal effects,

pe : vector of random permanent environmental effects and nonadditive genetic effects, and

e : vector of random residual effect.

X, Z, and W are incidence matrices relating records to fixed animal and permanent environmental effects, respectively.

Software and Analysis Application:

The BUGA software used in the research is programmed in PYTHON 3x (Anonymous, 2023c) programming language, with some parts also utilizing the C language. It is currently in an alpha version and is continuously being developed.

The BUGA software can estimate genetic and environmental effects, along with variance components, using the Restricted Maximum Log likelihood (REML) method and Best Linear Unbiased Prediction (BLUP). It employs two algorithms, Average Information (AI) and Expectation-Maximization (EM). Depending on the selected model, it can estimate phenotypic variance components (genetic, permanent environmental, and random environmental), as well as additive genetic and permanent environmental effects for animals.

The software, as seen in Figure 1, has a visual interface. In the top-left corner of the interface, there

are buttons for opening files, selecting the folder for saving analysis results, running the program, clearing the screen, and closing the program. On the side, there is a combo box that lists the models the user can choose, allowing them to select a model without the need to prepare a parameter file. On the left side of the interface, options for solving the relationship matrix (algebraic or Henderson methods), the output of this matrix, and the algorithm to be used in the solution (AIREML or EMREML) are provided as optional choices for the user.

The user can also optionally specify convergence criteria.

On the fixed effects page, the fixed effects in the model are coded and their names are entered in the header row. In the observation page, the repeated observations obtained for each animal are entered based on the animal number, and the name of the observation is written in the header row (In this study, as milk yield is considered, it is simply labeled as "süt_ver; milk_yield"). Finally, in the starting values page, initial values for genetic variance, permanent environmental variance, and random environmental variance are entered in cells B2, B3, and B4 (Figure 2).

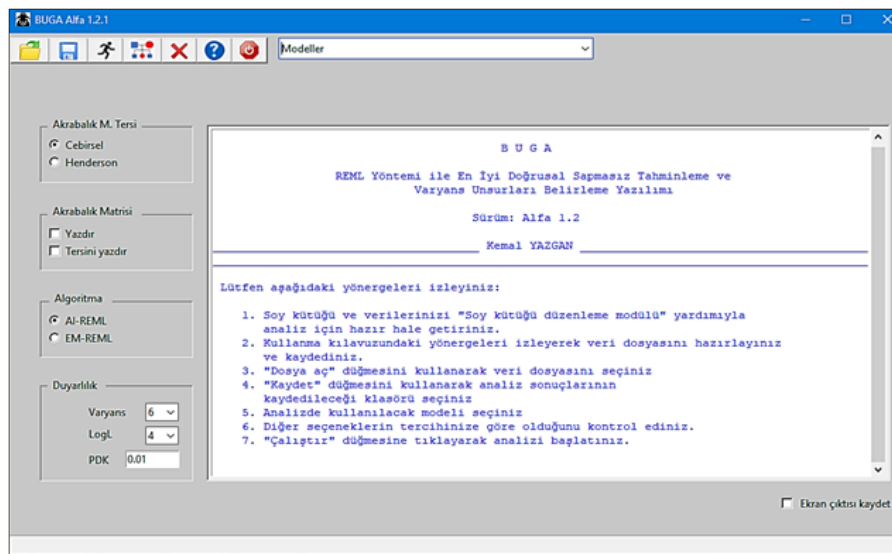


Figure 1. Interface of BUGA software

Default values are shown in the bottom-left corner of Figure 1 (Duyarlılık; Sensitivity). In addition, the software has its own screen for tracking the operations performed during runtime. Users can monitor these operations from the start of the program until completion, and they can choose to save them by checking the "Save Screen Output (Ekran çıktısı kaydet)" box in the bottom-right corner of the interface.

The software requires only one MSEXcel file for input (Figure 2). The pages of this file (for the model used in this study) should be named in the following order: pedigree, permanent environmental, fixed effects, observation, and starting values. The first rows on these pages are header rows. Subsequently, on the pedigree page, animal sire and dam numbers should be entered in sequence, starting from 1, with the header rows labelled as animal, sire, and dam, respectively.

Similarly, on the permanent environmental page, the name of the permanent environmental effect (In this study, a permanent environmental page has been created since repeated observations are used. This is not mandatory for a simple animal model) can be written in the first row. The permanent environmental effect contains the animal number for each animal as many times as it has repeated observations. For example, if animal number 25 has three observations, number 25 is written three times in succession.

After these steps, the data file is closed, and the user selects the location of this file and the folder where the output files will be saved using the relevant buttons on the interface. Afterward, the user selects the appropriate model (Equation 1), as shown in Figure 3, and clicks on the "running man" icon to execute the program.

After completing the software process, it displays the message "PROGRAM SONLANDI; PROGRAM ENDED" on its dedicated screen, indicating that the program has successfully concluded. If the user has made errors in preparing the data file or selected the wrong model, the software will identify where the error occurred and notify the user with an error code on its own screen. After completing its operation, the software generates five output files for this model, including two MSEXcel files and three graphs (Figures 4 - 8).

These files are named as follows: Pedigree_Solution.xlsx; Soy_Kütüğü_Çözümü.xlsx, Results.xlsx; Sonuçlar.xlsx, genetic_effects.jpg; genetik_etkiler.jpg, Permanent_Environmental_Effects.jpg; Kalıcı çevre etkileri.jpg and Kalıntılar.jpg; Residuals.jpg. In the Pedigree_Solution.xlsx output file, various statistical information related to the pedigree is provided (Figure 4). Additionally, if the user has requested the

The figure displays four screenshots of an Excel spreadsheet titled "Düzenlenmiş_veri2 - Excel". The screenshots show different sections of the data file prepared for BUGA software.

Top Left Screenshot (L21): Shows a table with columns A, B, C, and D. The data is as follows:

	A	B	C	D
1	hayvan	baba	ana	
350	349	0	7	
351	350	0	8	
352	351	0	112	
353	352	0	244	
354	353	0	9	
355	354	0	254	
356	355	0	146	
357	356	0	10	
358	357	0	256	
359	358	0	302	
360	359	0	11	
361	360	0	102	
362	361	0	127	
363	362	0	12	
364	363	0	0	
365	364	0	0	
366	365	0	0	
367	366	0	0	
368	367	0	295	
369	368	0	49	
370	369	0	207	

Top Right Screenshot (F13): Shows a table with columns A, B, C, D, E, F, G, H, I, J. The data is as follows:

	A	B	C	D	E	F	G	H	I	J
1	hayvan									
2	24									
3	24									
4	25									
5	25									
6	25									
7	26									
8	26									
9	26									
10	27									
11	27									
12	27									
13	27									
14	27									
15	27									
16	28									
17	29									
18	29									
19	29									
20	30									
21	30									
22	30									

Bottom Left Screenshot (G13): Shows a table with columns A, B, C, D, E, F, G, H. The data is as follows:

	A	B	C	D	E	F	G	H
1	lak_sir	malak_ay	lak_sur_kod					
2	1	6	7					
3	2	10	3					
4	1	1	7					
5	2	2	7					
6	3	4	7					
7	1	12	5					
8	2	2	5					
9	3	1	6					
10	1	6	3					
11	2	4	6					
12	3	5	2					
13	4	3	1					
14	5	3	5					
15	6	9	1					
16	1	6	6					
17	1	5	2					
18	2	4	2					
19	3	4	2					
20	2	5	6					
21	3	8	5					
22	4	8	6					

Bottom Right Screenshot (M14): Shows a table with columns A, B, C, D, E, F, G, H, I. The data is as follows:

	A	B	C	D	E	F	G	H	I
1	sut_ver								
2	1742								
3	1122								
4	2440								
5	2114								
6	2348								
7	1838								
8	1581								
9	2205								
10	1115.62								
11	798								
12	696								
13	570								
14	1116								
15	544.6								
16	1541.01								
17	882								
18	829								
19	917								
20	1498								
21	1493								
22	1653								

Bottom Center Screenshot (G8): Shows a table with columns A, B, C, D, E, F, G, H. The data is as follows:

	A	B	C	D	E	F	G	H
1	varyans	değer						
2	genetik	10						
3	kalcı çevre	100						
4	tesadüfî çevre	1000						
5								
6								
7								
8								
9								
10								

Figure 2: Data file and sections etklr prepared with MSEXcel for BUGA software

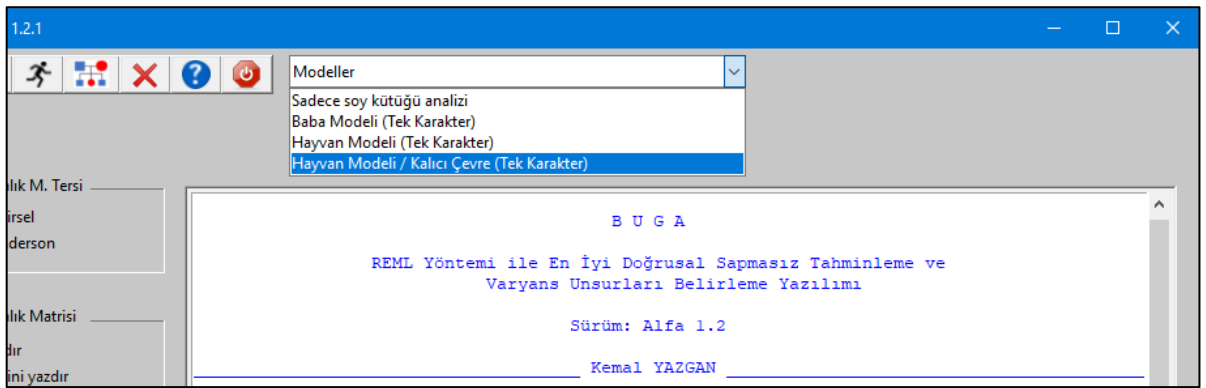


Figure 3. Model selection.

ÖZELLİK	DEĞERLER
Hayvan sayısı :	442
Baba sayısı :	0
Ana sayısı :	63
Anası ve babası bilinmeyen birey sayısı :	379
Sadece babası bilinmeyen birey sayısı :	63
Sadece anası bilinmeyen birey sayısı :	0
Anası ve babası bilinen birey sayısı :	0
Ebeveyn olmuş birey sayısı :	63
Ebeveyn olmuş birey sayısı (%) :	14,25339
Akrabalı yetişmiş birey sayısı :	0
Akrabalı yetişmiş birey sayısı (%) :	0
Ortalama akrabalı yetiştirme katsayısı :	0
Soy kütüğünde verim kaydı bulunmayan birey sayısı :	23
Soy kütüğünde verim kaydı bulunan birey sayısı :	419
Veri dosyasında verim kaydı olanların tüm hayvanlara oranı (%) :	94,79638

Figure 4. Soy_Kütüğü_Çözümü.xlsx (Pedigree_Solution.xlsx) file and its contents.

relationship matrix and/or its inverse, these are also included in different pages within the same file. The Results.xlsx file consists of an analysis report, fixed effect solutions, genetic effect solutions, permanent environmental solutions, genetic parameters, log-likelihood changes, and residuals pages (Figure 5).

In the genetic effect page, the estimated additive genetic effect for each animal is listed, along with standard errors for each estimate and the correlation values between the true and estimated values. These values can be considered as the breeding values of the animals. In addition, values related to inbreeding coefficients for each animal are provided. As shown in Figure 5, the standard errors of the estimated variance components are also given.

Result and Discussion

Upon completion of the program's operation, as can be seen from the output files it generated (Figure 5), the values for residual variance, genotypic variance, and permanent environmental variance were estimated as

107112.60 kg², 20045.22 kg², and 1259.50 kg², respectively. The heritability (h²) and repeatability (r) were determined as 0.1561 and 0.1659, respectively. Furthermore, it was observed that the additive genetic effects for the 442 animals ranged from +235.76 to -246.60 kg. In other words, the animal with the highest additive genetic effect produced 235.76 kg more milk than the population average. This value is solely attributable to the genetic effect of the animal, independent of environmental factors. Therefore, the additive genetic effect values for animals can be used as selection criteria.

When compared to other software programs, this software offers advantages such as the ability for the user to directly select the model they want to work with, the absence of the need to prepare parameter files, the ability to prepare data files in MSEXcel format and obtain result files in the same format, and the overall ease of use. This makes it possible for the software to become widespread and effectively used in field conditions.

Analiz raporu

Veri dosyası adı: C:/Users/Kemal YAZGAN/Desktop/soysal_saray/ddt/Düzenlenmiş_veri2.xlsx
 Analiz tipi: Hayvan Modeli/Kalıcı Çevre (Tek Karakter)
 Algoritma: AI-REML
 NRM Ters metodu: Cebirsel
 AIC değeri: -4704,9511
 BIC değeri: -4711,9967
 Log olasılık değeri: -4701,9511
 Karakter adı: sut_ver
 Maksimum: 3391
 Minimum: 51
 Ortalama: 1251,51
 Standart sapma: 523,086
 Standart hata: 18,391
 Varyasyon katsayısı(%): 41,8
 Tarih / Saat: 21.10.2023 / 17:24:33

Sabit Etki(ler)

Sabit Etki(ler)	Seviye	n	Ortalama	Standart hata	Çözüm	Etki miktarı
1 lak_sir	1	349	1246,69	29,689	61,1599801	-36,38691127
2 lak_sir	2	229	1260,89	33,552	70,41059603	-27,13629534
3 lak_sir	3	155	1250,58	40,784	124,0880242	26,54113281
4 lak_sir	4	55	1344,43	62,722	77,88646613	-19,66042523
5 lak_sir	5	15	968,1	74,612	47,31560546	-50,23128591
6 lak_sir	6	6	983,13	174,922	99,67953124	2,13269874
7 lak_sir	7	1	1682,85		202,2880364	104,7411451
8 malak_ay	1	37	1392,35	93,503	-291,0788248	92,92026799
9 malak_ay	2	46	1295,71	74,954	-332,4106617	51,58843109
10 malak_ay	3	67	1344,15	76,244	-350,6426628	33,35643001
11 malak_ay	4	76	1325,76	56,426	-360,615175	23,38391774
12 malak_ay	5	88	1382,79	50,537	-331,5778215	52,42127127
13 malak_ay	6	111	1385,44	48,295	-412,5804823	82,4186105
14 malak_ay	7	94	1253,98	51,245	-424,3186478	-28,16540428
15 malak_ay	8	76	1135,58	52,548	-424,3186478	-40,31955503
16 malak_ay	9	77	1041,46	54,861	-464,5164763	-80,51738352
17 malak_ay	10	55	1172,82	72,722	-431,8286093	-47,82951654
18 malak_ay	11	41	1031,35	72,931	-505,5725154	-122,5734226
19 malak_ay	12	42	1075,22	84,702	-400,6827394	-16,68364662
20 lak_sur_kod	1	83	623,89	27,537	960,7306114	-651,6966469
21 lak_sur_kod	2	105	804,87	23,804	1111,292667	-501,1345914

Genetik etki çözümleri

hayvan	Çözüm	Standart hata	Korelasyon	Akrabalı yetiştirme katsayısı
1	-22,52004631	138,8462627	0,19560293	1
2	-19,65085572	138,8707264	0,194734624	1
3	-7,659059857	138,8692534	0,19478702	1
4	-19,09189316	138,8812058	0,194361436	1
5	5,280549449	138,8590005	0,19515132	1
6	-5,369476241	138,8467203	0,195586724	1
7	-19,71265851	138,8345453	0,194242354	1
8	7,538857292	138,8490374	0,195504647	1
9	-31,65686902	138,8350819	0,195998445	1
10	-11,59388808	138,8498656	0,1954753	1
11	117,160737	138,8939144	0,193907863	1
12	5,733036607	138,8462627	0,19560293	1
13	43,20663096	138,8380535	0,195893406	1
14	67,97692663	138,8354408	0,195985764	1
15	-3,359306021	138,8798534	0,194409638	1
16	55,3654692	138,8676722	0,194843247	1
17	-24,72689115	138,845581	0,195627069	1
18	0,276914705	138,875062	0,194580315	1
19	-17,1145624	138,8484153	0,195526685	1
20	7,509566535	138,8707264	0,194734624	1
21	1,29700142	138,8402979	0,195814035	1

Genetik parametre

Parametre	Tahmin	Standart hata
2 Fenotipik varyans	128417,3397	-
3 Genotipik varyans	20045,22512	33090,13322
4 Kalıcı çevre varyansı	1259,508833	33442,74873
5 Teasdüflü çevre varyansı (Hata)	107112,6058	7514,428929
6 Kovaryans (a,e)	0	-
7 Kovaryans (pe,e)	0	-
8 Kovaryans (a,pe)	0	-
9 Kalıtım derecesi (h ²)	0,1561	0,21920448
10 Tekrarlanma derecesi(r)	0,1659	0,301899833
11 e ²	0,834097685	-

Figure 5. Sonuçlar.xlsx (Results.xlsx) file and its contents.

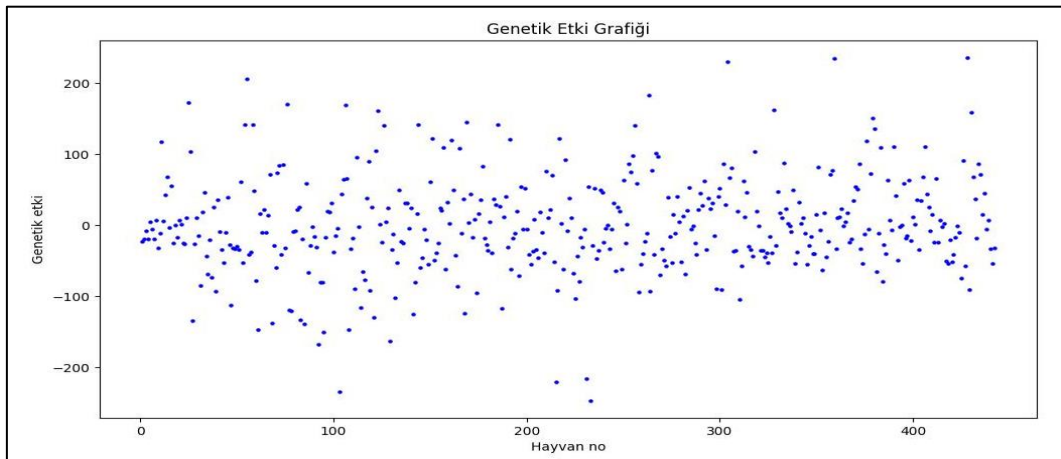


Figure 6. Scatter diagram of additive genetic effect plot.

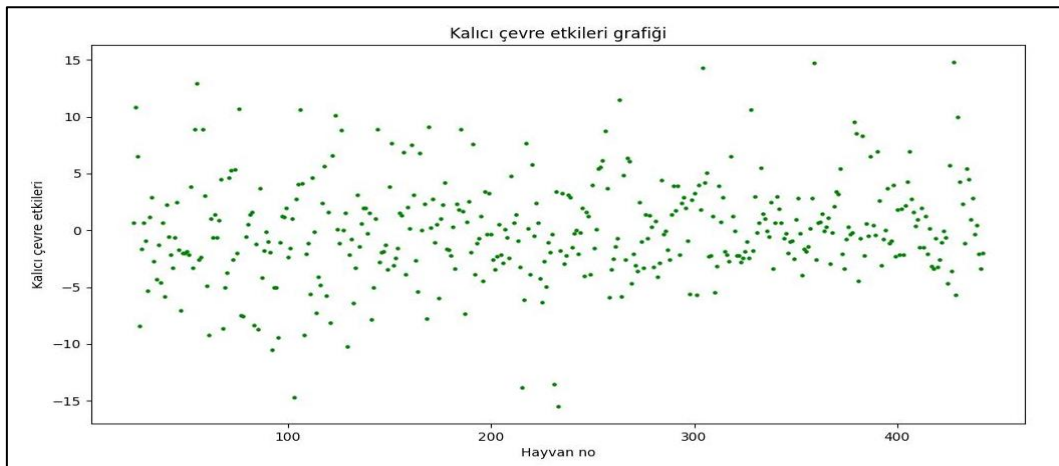


Figure 7. Scatter diagram of permanent environmental impacts plot.

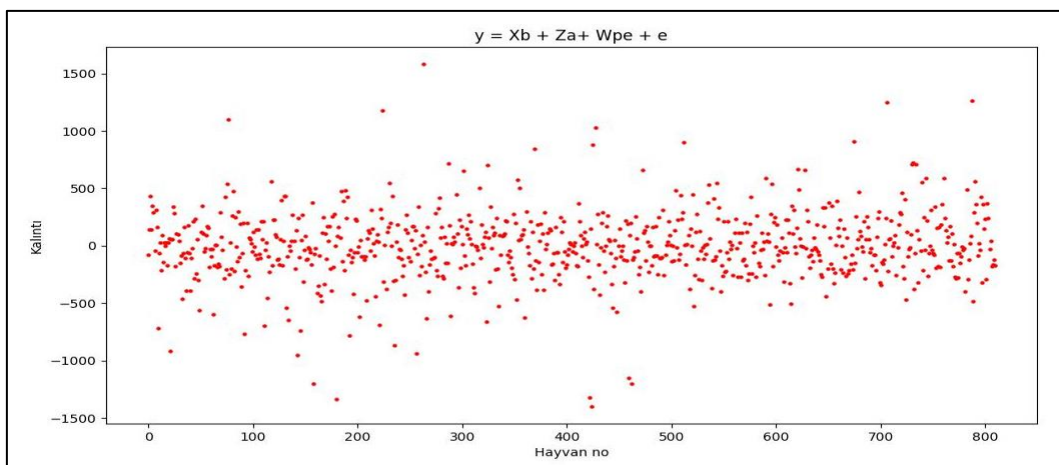


Figure 8. Scatter diagram of residuals (kalıntılar) plot.

Conclusion

The results obtained from this case study demonstrate that the BUGA software can be easily employed in farm recording conditions due to its user-friendly features. Furthermore, if the usage of this software with its unique user-friendly features is promoted, it is expected to make a significant contribution to the efficiency of breeding programs. This will increase the accuracy of selection studies in breeders' enterprises with a regular registration system. The software is continually being enhanced and configured to handle more complex models in the future.

Conflicts of Interest

There is no conflict of interest between the authors.

References

- Anonymous, 2023a. <http://nce.ads.uga.edu/wiki/doku.php> Access date: 02.11.2023
- Anonymous, 2023b. www.mixblup.eu/ Access date: 02.11.2023
- Anonymous, 2023c. <https://www.python.org/> Access date: 02.11.2023
- Boldmon, K. G., Kriese, L.A., Van Vleck, L.D., Van Tassel, C.P. ve Kachman, S.D. 1995. A manual for use of MTDFREML. A Set of Programs to Obtain Estimates of Variances and Covariances. US Department of Agriculture, Agriculture Research Service. USA
- Gilmour, A.R., Cullis, B.R., Welham, S.J., Thompson, R. 1998. ASREML User manual.
- Kumlu, S. 2003, Hayvan ıslahı. Türkiye damızlık sığır yetiştiricileri merkez birliğiyayınları, 296 sayfa, Ankara.
- Simm, G., Pollott, G., Mrode, R., Houston, R., Marshall, K. (2022) Çeviri Editörü: Mehmet İhsan Soysal Çiftlik Hayvanlarının Genetik İslahı ;Genetic Improvement of Farmed Animals .Nobel akademik yayıncılık ISBN978-625-417-903-7
- Meyer, K. 2007. "WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML)", J. Zhejiang Univ. - Sci. B 8, 815–821, <https://doi.org/10.1631/jzus.2007.B0815>
- Mrode, R. Linear models for the prediction of animal breeding values. (Second Edition), Cambridge, ma 02139 USA.
- SAS institute Inc. (2000). SAS User's guide statistics, version ed. SAS Institute, Gary. N.C. http://www2.sas.com/pdfs/s2k/v1_psm.pdf