

Applying a software to simulate phenotypic and pedigree information to improve swine genetic

**Wootichai Kenchaiwong¹, Monchai Duangjinda^{1,2*}, Wuttigrai Boonkum^{1,2}
and John Mabry³**

ABSTRACT: The application of the selection process through simulation had allowed farmers to study the test results, theoretical verification, and genetic evaluation of a proposed strategy in theory before the breeding program would be actually deployed. SIMF90P is a simulation program for validating the purpose of the most genetic response based on a terminal line index along with the feed conversion ratio (FCR), days to market weight (DAY) and percent lean (PCL). The empirical standard deviations of true breeding value (TBV) were closed to the predicted value which calculated by theoretical equation in all traits. The simulation program resulted high accuracy which composed of 0.72, 0.86, and 0.87 for PCL, FCR, and DAY respectively. The regression coefficient of genetic trend between average true and estimated breeding value (Avg-TBV and Avg-EBV) was similar for PCL and FCR and also nearly unbiased for DAY. In conclusion, this software could be applied to the swine genetic improvement to test the selection strategies.

Keywords: simulation program, feed conversion ratio, days to market weight, percent lean.

Introduction

Monte-Carlo simulation is a method of random number generation and probability density distribution, through the performance of multiple trial runs; based upon a mathematic model. This method has been applied within animal breeding schemes, generating performance data; based upon the mixed model generated for the animal (Duangjinda et al., 2001; Satoh et al., 2002). Model simulation is often used in investigative problems or in mating selection suitability because the actual testing of individual animals would require an enormously large test population in the best breeding strategies. Moreover, determining feed efficiency or carcass traits are generally very difficult, and expensive to obtain.

Modeling for animal selection and mating design have been previously investigated through simulation study (Muir, 1997; Pedersen et al., 2009).

Feed efficiency, growth rate, and percent lean are emphasized in selection programs within the pig industry. Although the feed conversion ratio (FCR) improved, the majority of such studies had been conducted with few actual measurements of feed intake (Kuhlers et al., 2003; Hoque et al., 2007). We presented SIMF90P: a software to simulate phenotypic and pedigree information to improve swine genetic, which is the application of the selection process through simulation would allow farmers to study the test results, theoretical verification, and genetic evaluation of a proposed strategy in theory; before the actual deployment

¹ Department of Animal Science, Faculty of Agriculture, Khon Kaen University, Khon Kaen 40002, Thailand.

² Research and Development Network Center for Animal Breeding, Khon Kaen University, Khon Kaen 40002, Thailand

³ Department of Animal Science, Iowa State University, Ames 50011.

* Corresponding author: monchai@kku.ac.th

of a breeding program. This software has a commercial benefit and could be applied for commercial pig industry. Therefore, the objective of this study is to examine the variables within the simulation software of such a genetic evaluation program.

Materials and Methods

Implementation

The Monte-Carlo simulation, based upon a problem solving method, generates suitable

$$BV_i \sim N(0, \sigma_a^2), Litter_i \sim N(0, \sigma_c^2) \text{ and } Err_i \sim N(0, \sigma_e^2) \tag{1}$$

Gender and contemporary groups (CG) were generated for the animals under descriptive uniform distribution ($X_i \sim U(a,b)$), as fixed effects. The X_i is variable in each fixed effect of animal i ;

random numbers, and records the results. The simulator, using Fortran 90, begins generating the true breeding value (TBV) and residual error of animals in base generation (G0), after which the genetic variance, common litter variance, and residual variance under normal distribution are determined; given the following equation:

and a,b is the parameter of uniform distribution in each fixed effect. The effect of gender and CG were estimated, based on the variance of each fixed effect under normal distribution; as:

$$(CG_n \sim N(0, \sigma_{cg}^2) \text{ and } Sex_{1,2} \sim N(0, \sigma_{sex}^2)) \tag{2}$$

Boars and gilts in G0 are randomly mated for producing the next generation population. After G0, the true breeding value of an animal was calculated, expressed in Equation (3); and the

fixed types and effects were similarly generated in G0 in every generation. The general observations for offspring (combining all effects), were determined using Equation (4).

$$BV_i = \frac{1}{2}BV_s + \frac{1}{2}BV_d + \varphi \tag{3}$$

$$y_{ij} = \mu + Cg_{ij} + Sex_{ij} + BV_{ij} + Litter_{ij} + Err_{ij} \tag{4}$$

Where y_{ij} is phenotype of animal i , and trait j , μ is overall mean, BV_i , BV_s , and BV_d are the breeding values of animal i (sire and dam); and φ is the Mendelian sampling variance, generated under $\varphi_i \sim N(0, \frac{1}{2}\sigma_a^2)$, Cg_{ij} is contemporary groups effect, Sex_{ij} is gender effect, and $Litter_{ij}$ is common litter of animal i , and trait j .

Selection and mating

The data, generated by simulation, includes all primary information, were shown in **Table 1**. In this study, FCR applies to an entire life time, PCL were adjusted to approximately 105kg of days to market weight (DAY). Categorically, there were two genders, and five contemporary groups (CG)

per generation. The base population of 50 boars and 150 gilts (population size = 150) were used for the G₀, which was considered randomized mating. The numbers born alive were generated among 8 piglets per litter, with a standard deviation of ± 3 . The production of ~ 128 litters, after considering a 15% fallout rate; was generated an approximate 50:50 ratio of males and females within this study. The next 50 boars

and 150 gilts of the previous generations were selected and mated to produce the offspring of next generation. The estimated breeding values (EBV) were estimated through BLUP in selecting animals along the terminal line index (TLI), applying Equation (5). The economic value in each trait was estimated by multiple regression method (Van Vleck, 1993).

$$TLI = 100 - 30 * EBV_{(FCR)} + 2.0 * EBV_{(PLC)} - 1.75 * EBV_{(DAY)} \quad (5)$$

Program testing

In the last generation (G₆), genetic parameters were estimated once again, under multiple trait analysis with each animal model, similar to equation (3). The average estimated breeding values (EBV) in each trait, within each generation; were used for measuring genetic responses. The overall genetic response in each generation was estimated as a linear slope regression of responses. In our study, thirty replications were conducted. Program optimizations were based upon:

a) Correlation between TBV and EBV for each trait. This should be equal to the 'accuracy' of each trait. Again, we looked to see if this changed over generations, or if it matches what was expected.

b) Standard deviation of the phenotypes for each trait, which determined any changes over a given generation for each trait.

c) Standard deviation of the TBV (σ_{TBV}) for each trait, which determined any changes over a given generation and closed to the predicted values ($\hat{\sigma}_{TBV}$).

d) An accordance of true response and estimated response. Estimates of variance components and corresponding values were averaged from 30 replications.

Results and discussion

The base population contained of 50 sires and 150 dams. The pedigree data of animals in over six generations; consisted of approximately 6,065 pigs, which was in accordance with the 15% fallout rate and number born alive; generating litters of 8 (± 3) piglets. The approximate variances for gender were fixed in each generation, at 22%, 18%, and 26% in all random variances; and 11%, 12%, and 17% in the CG variances for FCR, PCL, and DAY, respectively.

Description of simulated data

The normal distribution of phenotypic data was tested; using a normal probability plot, which generated data of phenotypes over six generations. Distribution was considered normal, and emptied for all outliers (data not shown). The phenotype means decreased for DAY and FCR, based upon

the selection; which increased slightly in PCL (Table 2). The regression of the phenotypes in each generation decreased for downward selection traits (-4.87 days/generation for DAY, and -0.11 kg feed/gain/generation for FCR); and increased for upward selection traits (0.18 percent/generation for PCL). These results were in accordance with the reality selection projection, as well as the results of previous studies, having a similar selection process (Chen et al., 2003; Kuhlert et al., 2003; Nguyen and McPhee, 2005). The deviation of means varied in each generation, yet was close to the phenotypic standard deviation plugs for FCR, PCL, and DAY; at 0.30, 3.10, and 14.14, respectively. The minimum and maximum values of each trait, in each generation, were within the range (with a ± 3 standard deviation) of approximately 99% of our distribution. Our study configured the software cloud required to test each selection strategy, within the genetic evaluation program.

Variance components of six generations are shown in Table 3. The Variance components and heritability of traits therein were close to the primary information (shown in Table 1), which was consistent with previous studies, based upon reality selection (Nguyen and McPhee, 2005;

Hoque et al., 2007; Hoque and Suzuki, 2008). This study has determined that the simulated software proved compatible with general principles of genetic variance and realized heritabilities.

Correctness of the simulation program

The overall means and standard deviation of the phenotypes, and true breeding values of each trait changed over six generations (Table 4). This may be explained by the genetic response improvement, based on selection; presented in the equation $R = i * h^2 * \sigma_p$, where R is the genetic response, i is selection intensity, and σ_p is the standard deviation of the phenotype (Falconer and Mackay 1996). The CV percentage for FCR, PCL, and DAY were minimal, at 13.82%, 5.79%, and 11.18%, respectively. In a past study, CV for growth and composition traits were measured at 11% to 29%. The CV percentages of FCR and PCL within the selection line closely reflected the report by Nguyen and McPhee (2005). Likewise, Holm et al. (2004); reported CV percentages for DAY (at 100 kg) and PCL at 10.69% and 5.04%, respectively; in Norwegian Landrace pigs. Our simulations parallel those within the reality of selection, theorized in Equation (6) as:

$$Var(EBV) = (r_{TBV,EBV})^2 * Var(TBV) \quad (6)$$

where, the variance of TBV ($Var(TBV)$) is equal to V_A . From this, we may reversibly predict the standard deviation of TBV, as expressed in Equation: $\sigma_{TBV} = \sigma_{EBV} / r_{TBV,EBV}$, where $r_{TBV,EBV}$ is the correlation between TBV and EBV. In this study, the predictions of σ_{TBV} were 0.24, 1.82, and 11.07

for FCR, PCL, and DAY, respectively. We found that the empirical standard deviation of TBV was close to the predicted values of all traits within our study. The data simulation has therefore proven to be in agreement with the predicted standard deviations of TBV.

Accuracy and response to selection

The correlation between true breeding value and estimated breeding value ($r_{TBV,EBV}$) for each trait is defined by the accuracy of each trait (Bijma, 2012). In genetic evaluation, the accuracy can be calculated from prediction error variance (PEV) and additive genetic variance, followed as equation; $r = \sqrt{1 - PEV/\sigma_a^2}$ (Miszta and Wiggans, 1988). Bijma (2012) demonstrated that there are conceptually different things between $r_{TBV,EBV}$ and accuracy from genetic evaluation. The selection response was reflected from $r_{TBV,EBV}$ but does not the prediction error variance (PEV), which is consistent with our study for PCL. The accuracy form $r_{TBV,EBV}$ for PCL in our study was approximately 0.72. Therefore, the acceleration of genetic response was considerably small. While, the simulation program can be give high accuracy from $r_{TBV,EBV}$ approximately 86% for FCR and 0.87% for DAY.

The patterns of the genetic trends of selection are shown in **Figure 1**. The genetic trends in all traits were consistent with the responses found

within the phenotypic trends. The average EBV and TBV in each generation did not differ significantly in FCR and DAY; but differed conversely in the empirical trends of PCL. This is in agreement with the higher accuracy of FCR and DAY, whereas the selection accuracy of PCL was lowest, when compared with other traits (**Table 4**). The accuracy of PCL in this study (<75%), yielded a high differential percentage of average TBV and EBV responses in each generation (34-56% during the first six generations), which implies the further need for theoretical verification. The results further indicated that the differential percentages of the regression coefficient of genetic trends between average EBV and TBV were similar in PCL (0.18% per generation), as well as in the regression of FCR (-0.11 kg of feed/kg of gain, in each generation). The regressions of the average EBV and TBV for DAY were between -5.09 and -4.93 days per generation, and were considered unbiased within the evaluation.

Table 1 Parameters used in simulation to generate phenotypic traits, contemporary group variance (σ_{CG}^2) sex variance (σ_{sex}^2), additive variance (σ_a^2), common litter variance (σ_c^2), error variance (σ_e^2) and genetic correlation (r_g).

Traits ^{1/}	Mean	σ_{CG}^2	σ_{sex}^2	σ_a^2	σ_c^2	σ_e^2	h^2	c^2	r_g	
									PCL	DAY
FCR	3.000	0.010	0.017	0.028	0.005	0.050	0.31	0.06	-0.29	0.30
PCL (%)	56.00	1.17	1.76	3.36	0.38	5.86	0.35	0.04	-	0.30
DAY (days)	170.00	35.00	52.00	56.00	40.00	103.00	0.28	0.20	-	-

^{1/} FCR= feed conversion ratio, PCL= percentage lean content (%), and DAY = day to market weight (days).

Table 2 Descriptive analysis of phenotypes through simulation over six generations, indicating days to market weight (DAY), percent lean (PCL), and feed conversion ratio (FCR); based on random mating (applying a mean of 30 replicate simulations).

Generation	DAY			PCL			FCR		
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max
1	167.30 (15.08)	117.50	220.75	55.47 (3.24)	45.52	66.65	3.04 (0.36)	1.77	4.38
2	167.74 (15.46)	121.31	215.37	55.87 (3.11)	45.14	65.02	2.89 (0.33)	1.78	4.36
3	160.98 (16.30)	110.99	209.70	55.59 (3.29)	44.30	66.20	2.82 (0.32)	1.65	3.83
4	152.76 (16.11)	102.38	202.37	56.09 (3.12)	46.12	65.59	2.71 (0.34)	1.64	3.73
5	151.49 (14.93)	104.41	197.82	56.33 (3.39)	45.58	67.20	2.61 (0.33)	1.56	3.64
6	144.64 (14.88)	93.14	190.12	56.34 (3.33)	44.37	66.63	2.49 (0.33)	1.42	3.62

Table 3 Average variance components, genetic parameters, and empirical standard error under mate selection involving six generations (average estimates over 30 replications).

Traits ^{1/}	Genetic parameters ^{2/}						r_g ^{3/}		
	σ_a^2	σ_c^2	σ_e^2	σ_p^2	h^2	c^2	FCR	PCL	DAY
FCR	0.030 (0.001)	0.005 (0.000)	0.056 (0.001)	0.090 (0.002)	0.33 (0.005)	0.05 (0.001)	-	-0.32	0.30
PCL	3.24 (0.056)	0.36 (0.006)	5.84 (0.107)	9.43 (0.17)	0.34 (0.006)	0.04 (0.001)		-	0.26
DAY	56.38 (1.02)	39.43 (0.71)	102.97 (1.89)	198.78 (3.60)	0.28 (0.004)	0.20 (0.003)			-

^{1/} Traits: FCR = feed conversion ratio, PCL = percent lean (%), DAY = day to market (days).

^{2/} Variance component (σ_a^2 = additive variance, σ_c^2 = common litters variance, and σ_p^2 = phenotypic variance) and numbers in parenthesis are standard errors.

^{3/} r_g ; the genetic correlation was calculated by genetic variance and covariance of the two traits.

Table 4 The overall means, standard deviations (SD), and tested correlations of phenotypic and true breeding values.

Traits	Phenotypic value		True breeding value			Parameter estimate ^{1/}	
	Overall mean	SD	Overall means	σ_{TBV}	$\hat{\sigma}_{TBV}$	$r_{Phe,TBV}$	$r_{TBV,EBV}$
FCR	2.75	0.38	-0.17	0.24	0.24	0.64	0.87
PCL	55.97	3.24	-0.06	1.82	1.82	0.55	0.72
DAY	157.91	17.66	-7.54	10.92	11.04	0.61	0.86

^{1/} $r_{Phe,TBV}$ is correlation between phenotype and true breeding value, and $r_{TBV,EBV}$ is correlation between true breeding value and estimated breeding value, which refer to accuracy of selection.

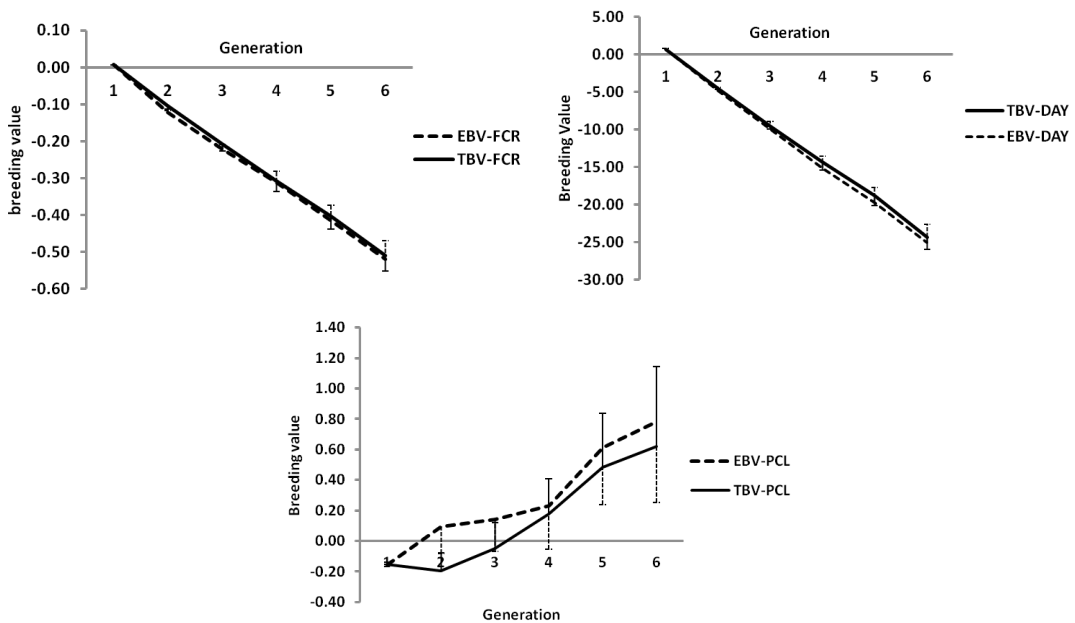


Figure 1 Trends in the average true and estimate breeding values of feed conversion ratio (FCR), percent lean (PCL), and days to market weight (DAY); based on positive assortative mating, given the mean and standard deviation of 30 replicate simulations.

Conclusions

The approximated minimum and maximum value of all traits are in ranged of three standard deviations. The empirical standard deviations of TBV were closed to the predicted value which calculated by theoretical equation in all traits.

They are also accorded to genetic variance and realized heritability under the mate selection. Correlation between true breeding value and estimated breeding value $90 (r_{T,I})$ for each trait is defined as the accuracy for each trait. The accuracies for each trait are changes over generations. The simulation program can be given high accuracy which composed of

0.72, 0.86, and 0.87 for PCL, FCR, and DAY respectively. The regression coefficient of genetic trend between Avg-EBV and Avg-TBV was similar for PCL and FCR and also nearly unbiased for DAY. This shows that the simulation software gives the consistent result with theoretical expectations and several literatures in reality. This study has shown that the software could be applied to the swine genetic evaluation to test the selection strategies.

Acknowledgements

This paper was partially funded by The Thailand Research Fund under Royal Golden Jubilee PhD. Program. The authors also thank the Research and Development Network Center for Animal Breeding, KKU-TRF for manuscript preparation.

References

- Bijma, P. 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. *J. Anim. Breed. Genet.* 129: 345-358.
- Bunter, K. L., W. Cai, D. J. Johnston, and J. C. M. Dekkers. 2010. Selection to reduce residual feed intake in pigs produces a correlated response in juvenile insulin-like growth factor-I concentration. *J. Anim. Sci.* 88: 1973-1981.
- Chen, P., T. J. Baas, J. C. M. Dekkers, K. J. Koehler, and J. W. Mabry. 2003. Evaluation of strategies for selection for lean growth rate in pigs. *J. Anim. Sci.* 81: 1150-1157.
- Duangjinda, M., I. Misztal, and J. K. Bertrand. 2001. Generating data in models including direct and maternal dominance effects. *J. Appl. Genet.* 42: 193-203.
- Falconer, D., and T. Mackay. 1996. *Introduction to Quantitative Genetics* 4th ed., Longman, Harlow, U.K.
- Holm, B., M. Bakken, G. Klemetsdal, and O. Vangen. 2004. Genetic correlations between reproduction and production traits in swine. *J. Anim. Sci.* 82: 3458-3464.
- Hoque, M. A., and K. Suzuki. 2008. Genetic parameters for production traits and measures of residual feed intake in Duroc and Landrace pigs. *Anim. Sci. J.* 79: 543-549.
- Hoque, M. A., K. Suzuki, H. Kadowaki, T. Shibata, and T. Oikawa. 2007. Genetic parameters for feed efficiency traits and their relationships with growth and carcass traits in Duroc pigs. *J. Anim. Breed. Genet.* 124: 108-116.
- Kuhlers, D. L., K. Nadarajah, S. B. Jungst, B. L. Anderson, and B. E. Gamble. 2003. Genetic selection for lean feed conversion in a closed line of Duroc pigs. *Livest. Prod. Sci.* 84: 75-82.
- Misztal, I., and G. R. Wiggans. 1988. Approximation of Prediction Error Variance in Large-Scale Animal Models. *J. Dairy Sci.* 71: 27-32.
- Muir, W. 1997. Genetic selection strategies: computer modeling. *Poult. Sci.* 76: 1066-1070.
- Nguyen, N., and C. McPhee. 2005. Genetic parameters and responses of performance and body composition traits in pigs selected for high and low growth rate on a fixed ration over a set time. *Genet. Sel. Evol.* 37: 199 - 213.
- Pedersen, L. D., A. C. Sørensen, M. Henryon, S. Ansari-Mahyari, and P. Berg. 2009. ADAM: A computer program to simulate selective breeding schemes for animals. *Livest. Sci.* 121: 343-344.
- Satoh, M., C. Hicks, K. Ishii, and T. Furukawa. 2002. Choice of statistical model for estimating genetic parameters using restricted maximum likelihood in swine. *J. Anim. Breed. Genet.* 119: 285-296.
- Van Vleck, L.D. 1993. *Selection index and introduction to mixed model methods.* CRC Press, Boca Raton, FL.