

162: Comparison of Genomic-Polygenic Evaluations Using Random Regression Models with Legendre Polynomials and Splines for Milk Yield and Fat Percentage in Thai Multibreed Dairy Cattle

Danai Jattawa¹, Skorn Koonawootrittriron¹, Mauricio A. Elzo², and Thanathip Suwanasopee¹

Kasetsart University, Chatuchak, Bangkok 10900, Thailand¹

University of Florida, Gainesville, FL 32611-0910, USA²

ABSTRACT

The objective of this research was to compare random regression models with third-order Legendre polynomials (RRLP) and linear splines with four knots (RRSP4) for milk yield (MY) and fat percentage (FP) genomic-polygenic evaluations in a Thai multibreed dairy cattle population. Models were compared using estimates of variance components, genetic parameters, model-fit criteria, accuracy of genomic-polygenic EBV (GPEBV), and animal rankings. The dataset contained pedigree and first-lactation monthly test-day records (69,029 for MY; 29,878 for FP) of 7,206 cows from 761 farms. Genotypic information included 74,144 actual and imputed SNP markers from 2,661 animals. Variance components and genetic parameters for MY and FP were estimated using REML procedures. The single-step random regression genomic-polygenic models included contemporary group (herd-year-season), calving age, heterozygosity, and population lactation curve regression coefficients as fixed effects. Random effects were animal additive genetic random regression coefficients, permanent environment random regression coefficients, and residual. The effects of population lactation curve, animal additive genetic, and permanent environment were fitted using regression coefficients of third-order Legendre polynomials for **RRLP** and splines with four knots for **RRSP4**. Results showed estimates of 305-day additive genetic variances and heritabilities for MY and FP to be higher for **RRLP** (additive genetic variances: 279,893.2 kg² for MY, 0.10 %² for FP; heritabilities: 0.27 for MY, 0.16 for FP) than for **RRSP4** (additive genetic: 260,178.1 kg² for MY, 0.08 %² for FP; heritabilities: 0.19 for MY, 0.11 for FP). Similarly, **RRLP** yielded a better model-fit and higher prediction accuracies than **RRSP4**. The **RRLP** had lower values of -2 log-likelihood (-2logL; 293,813), Akaike's information criteria (AIC; 293,855) and Bayesian information criteria (BIC; 293,915) as well as higher prediction accuracies (49.25% for MY; 38.63% for FP) than **RRSP4** (-2logL = 362,738; AIC = 362,888; BIC = 363,101; prediction accuracies: 47.23% for MY, 37.76% for FP). However, rank correlations between GPEBV from **RRLP** and **RRSP4** were high for MY (0.94) and FP (0.90). *Results indicated that GPEBV from RRLP should be preferred to those from RRSP4 to help increase selection responses for MY and FP in the Thai multibreed dairy cattle population.*

INTRODUCTION

Thailand initiated a project to develop a national Thai dairy genomic-polygenic evaluation system to improve the accuracy of animal selection in 2012. As a result, a single-step genomic-polygenic evaluation program was implemented in 2015 using a standard cumulative 305-d model (pedigree and phenotypes from 8,361 cows and genotypes from 2,661 animals with 74,144 SNP). This evaluation yielded accuracies that were on the average 7.2% higher than those from a conventional polygenic evaluation. In addition, genomic-polygenic evaluations in this population were also carried out using a random regression single-step model with third-order Legendre polynomials. This model yielded even more accurate genomic-polygenic EBV (GPEBV; 9% higher on the average) than those from the standard cumulative 305-d model. However, research has shown that splines would likely be better than Legendre polynomials for analyzing test-day phenotypes in random regression models (Misztal et al., 2006; Bohmanova et al., 2008). Use of splines was expected to further increase the accuracy of random regression genomic-polygenic evaluation in Thai dairy population. *Thus, the objectives of this research were to compare variance components, genetic parameters, model-fit criteria, accuracy of GPEBV, and animal rankings for milk yield (MY) and fat percentage (FP) using genomic-polygenic random regression models with third-order Legendre polynomials (RRLP) and splines with four knots (RRSP4) in a Thai multibreed dairy population.*

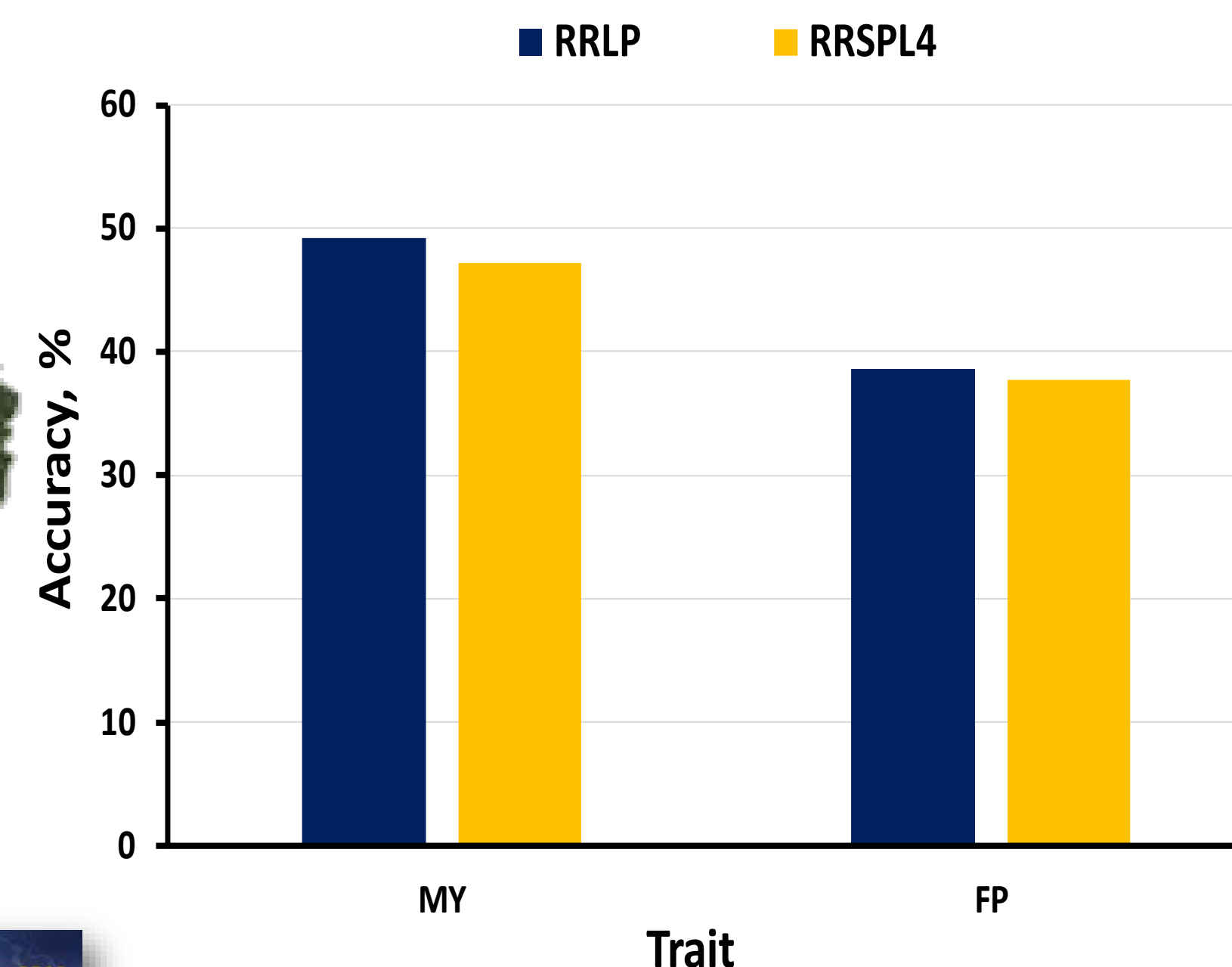


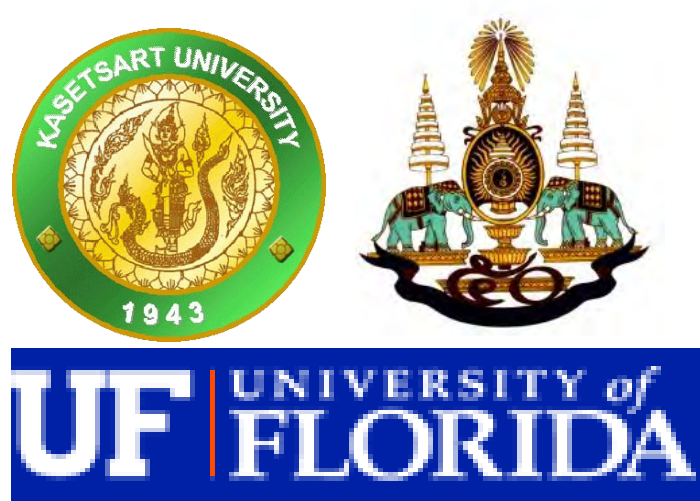
Fig. 1. Accuracy of GPEBV for milk yield (MY) and percentage (FP) using random regression models with third-order Legendre polynomials (**RRLP**) and linear splines with four knots (**RRSP4**)

MATERIALS AND METHODS

Animals, phenotypes, and genotypes: Pedigree and monthly test-day records of 69,029 milk yield (MY) and 29,878 fat percentages (FP) from 7,206 cows (716 farms) were used in this research. Cows were the progeny of 933 sires and 6,145 dams and had their first calving between 1997 and 2014. Tissue samples from 2,661 animals were collected and DNA extracted. Subsequently, DNA samples were genotyped using GeneSeek GGP9K (n = 1,412), GGP20K (n = 570), GGP26K (n = 540), and 80K (n = 139) chips. Animals genotyped with GGP9K, GGP20K, GGP26K chips were imputed to GGP80K using FImpute 2.2 (Sargolzaei et al., 2014). Actual and imputed SNP markers with minor allele frequencies lower than 0.04 (n = 2,375) and call rates lower than 0.9 (n = 175) were discarded. Consequently, the final genotype file contained 2,661 animals with 74,144 actual and imputed SNP markers.

Variance and covariance components: The estimates variance components, genetic parameters and GPEBV for MY and FP were obtained using single-step random regression genomic-polygenic models with third-order Legendre polynomials (RRLP) and linear splines with four knots (RRSP4). Knots were located at day 5, 65, 245, and 365 of lactation. The RRLP and RRSP4 models included contemporary group (herd-year-season), calving age, heterozygosity, and population lactation curve regression coefficients as fixed effects. Random effects were animal additive genetic random regression coefficients, permanent environmental random regression coefficients, and residual. The variance-covariance matrix of residual effects was assumed to be homogenous because of the small size of the dataset. Variance-covariance components for MY and FP were estimated using an average information restricted maximum likelihood algorithm (program AIREMLF90; Tsuruta, 2014). The computations of 305-day variance components and genetic parameters as well as GPEBV were described in detail in Jattawa et al. (2016).

Model comparisons: The comparisons between models were carried out based on the values of -2 log-likelihood (-2logL), Akaike's information criterion (AIC; Akaike, 1973), Bayesian information criterion (BIC; Schwarz, 1978), and Prediction accuracies. The prediction accuracies for each GPEBV was obtained as $\sqrt{1 - \frac{PEV_{a305d}}{\hat{\sigma}_{a305d}^2}} \times 100$, where PEV_{a305d} = the 305-day prediction error variance, and $\hat{\sigma}_{a305d}^2$ = the estimate of the 305-day additive genetic variance. In addition, rankings of animal GPEBV for MY and FY between RRLP and RRSP4 were also compared using Spearman's rank correlations.



162: Comparison of Genomic-Polygenic Evaluations Using Random Regression Models with Legendre Polynomials and Splines for Milk Yield and Fat Percentage in Thai Multibreed Dairy Cattle

Danai Jattawa¹, Skorn Koonawootrittriron¹, Mauricio A. Elzo², and Thanathip Suwanasopee¹



Kasetsart University, Chatuchak, Bangkok 10900, Thailand¹
University of Florida, Gainesville, FL 32611-0910, USA²

RESULTS AND DISCUSSION

Variance components and genetic parameters: Table 1 shows that additive genetic variances and covariances estimated with **RRLP** [var(MY): 279,893.2 kg²; var(FP): 0.10 %²; cov(MY-FP): -41.30 kg%] were slightly larger than those from **RRSP4** [var(MY): 260,178.1 kg²; var(FP): 0.08 %²; cov(MY-FP): -39.37 kg%]. Similarly, permanent environmental variances and covariances from **RRLP** [var(MY): 556,455.4 kg²; var(FP): 0.10 %²; cov(MY-FP): -72.90 kg%] tended to be larger than those from **RRSP4** [var(MY): 522,357.0 kg²; var(FP): 0.11 %²; cov(MY-FP): -58.64 kg%]. However, phenotypic variances and covariances from **RRLP** [var(MY): 1,023,747.6 kg²; var(FP): 0.66 %²; cov(MY-FP): -114.10 kg%] were somewhat lower than those from **RRSP4** [var(MY): 1,405,063.6 kg²; var(FP): 0.73 %²; cov(MY-FP): -98.29 kg%].

The lower phenotypic variances and covariances for **RRLP** than for **RRSP4** are largely explained by low **RRLP** residual variances and covariances. The residual variances and covariances from **RRLP** [var(MY): 187,399.0 kg²; var(FP): 0.46 %²; cov(MY-FP): -14.30 kg%] were substantially lower than those from **RRSP4** [var(MY): 622,528.5 kg²; var(FP): 0.54 %²; cov(MY-FP): -86.10 kg%]. *The **RRSP4** model may have underestimated additive genetic and permanent environmental variances and covariances. Perhaps fractions of additive genetic and permanent environmental variation unaccounted for by **RRSP4** were taken into the residual component resulting in overestimation of residual variances and covariances.*

The slightly higher additive genetic (4%) but somewhat lower phenotypic variances (21%) for MY and FP from **RRLP** than from **RRSP4** resulted in higher **RRLP** heritabilities (0.27 for MY; 0.16 for FP) than **RRSP4** heritabilities (0.19 for MY; 0.11 for; Table 2). The correlations between MY and FP obtained from **RRLP** and **RRSP4** were low and negative for additive genetic, permanent and temporary environment, and phenotypic effects. The higher estimates of variance components and genetic parameters for milk yield and fat percentage obtained with **RRLP** than with **RRSP4** indicated that Legendre polynomials accounted for genetic variation better than splines for these traits in the Thai dairy multibreed population. *Higher genetic parameter values from **RRLP** would be expected to yield more accurate GPEBV than with **RRSP4**, thus increasing the rate of genetic progress for milk yield and fat percentage in the Thai multibreed dairy population.*

Table 1. 305-day additive genetic, permanent environmental, phenotypic variances and covariances for milk yield (MY) and fat percentage (FP) estimated using two genomic-polygenic random regression models

Variance components	Model	
	RRLP	RRSP4
Additive genetic		
Var (MY), kg ²	279,893.2	260,178.1
Cov (MY, FP), kg%	-41.30	-39.37
Var (FP), % ²	0.10	0.08
Permanent environment		
Var (MY), kg ²	556,455.4	522,357.0
Cov (MY, FP), kg%	-72.90	-58.64
Var (FP), % ²	0.10	0.11
Temporary environment (residual)		
Var (MY), kg ²	187,399.0	622,528.5
Cov (MY, FP), kg%	-14.30	-86.10
Var (FP), % ²	0.46	0.54
Phenotypic		
Var (MY), kg ²	1,023,747.6	1,405,063.6
Cov (MY, FP), kg%	-114.10	-98.29
Var (FP), % ²	0.66	0.73

Model comparison: Table 3 shows model-fit criteria obtained with genomic-polygenic evaluations using **RRLP** and **RRSP4** models. The **RRLP** model yielded lower values of -2logL (293,813), AIC (293,855) and BIC (293,915) than **RRSP4** (-2logL: 362,738; AIC: 362,888; BIC: 363,101). Lower values for **RRLP** indicated that third-order Legendre polynomials provided a better fit to test-day milk yield and fat percentage lactation data than splines with four knots in the Thai dairy multibreed population. Similarly, model comparison based on GPEBV accuracies also showed that **RRLP** was more favorable than **RRSP4** for MY and FP (Fig. 1). Prediction accuracies obtained from **RRLP** (49.3% for MY; 38.6% for FP) were higher than from **RRSP4** (47.2% for MY; 37.8% for FP). Lastly, rankings between animal GPEBV from **RRLP** and **RRSP4** were highly correlated for MY (0.94) and FP (0.90; Table 4). *Genetic parameters, model fit criteria, GPEBV accuracies, and animal rankings obtained here indicated that GPEBV from **RRLP** should be preferred to those from **RRSP4** to help increase selection responses for MY and FP in the Thai multibreed dairy cattle population.*

Table 2. 305-day heritabilities and correlations for milk yield (MY) and fat percentage (FP) estimated using two genomic-polygenic random regression models

Parameters	Model	
	RRPL	RRSP4
Heritability (MY)	0.27	0.19
Heritability (FP)	0.16	0.11
Additive genetic correlation (MY, FP)	-0.24	-0.27
Permanent environmental correlation (MY, FP)	-0.31	-0.25
Phenotypic correlation (MY, FP)	-0.14	-0.10

Table 3. Estimates of -2 log-likelihood (-2logL), Akaike’s information criteria (AIC), and Bayesian information criteria (BIC) obtained using two genomic-polygenic random regression models

Model	Model fit criteria		
	-2logL	AIC	BIC
RRLP	293,813	293,855	293,915
RRSP4	362,738	362,888	363,101

Table 4. Rank correlations between animal GPEBV for 305-d milk yield (MY) and fat percentage (FP) evaluated using two genomic-polygenic random regression models

Trait	Rank correlation (RRLP, RRSP4)
MY	0.94
FP	0.90

FINAL REMARKS

- ❖ *Estimates of 305-day additive genetic variances and heritabilities for MY and FP to be higher for **RRLP** than for **RRSP4***
- ❖ ***RRLP** yielded a better model-fit and higher prediction accuracies than **RRSP4***
- ❖ *GPEBV from **RRLP** should be preferred to those from **RRSP4** to speed up selection response in the multibreed dairy cattle population*

REFERENCES

Akaike, H., 1973. Proc. 2nd Int. Symp. on Information Theory; Bohmanova et al., 2008. J. Dairy Sci. 91: 3627-3638; Jattawa et al. 2016. Livest. Sci. 188: 133-141; Misztal et al., 2006. J. Anim. Breed. Gent. 123: 74-80; Sargolzaei, et al., 2014. BMC Genomics, 15, 478; Schwarz, G., 1978. Ann. Stat. 6: 461-464; Tsuruta, S., 2014. <http://nce.ads.uga.edu>