



**Full Length Article**

# Estimation of Genetic Parameters and Breeding Values of Milk Yield Using Test Day Model with Cubic Splines and Legendre Polynomials in Chinese Holstein Cattle

Wan Lv<sup>1,2†</sup>, Gaozhan Cai<sup>1†</sup>, Yujiao Wang<sup>2</sup>, Xianguo Gu<sup>3</sup>, Xiuxin Zhao<sup>1</sup>, Xiaotao Liu<sup>3</sup>, Rongling Li<sup>1</sup>, Minghai Hou<sup>1</sup> and Jianbin Li<sup>1\*</sup>

<sup>1</sup>Dairy Cattle Research Center of Shandong Academy of Agricultural Sciences, Jinan 250100, China

<sup>2</sup>Department of College of Life Sciences and Food Engineering, Hebei University of Engineering, Handan 056001, China

<sup>3</sup>Ningbo Milk Group Co., Ltd, Ningbo, 315033, China

\*For correspondence: msdljb@163.com

Received 01 September 2020; Accepted 26 September 2020; Published 10 January 2021

<sup>†</sup>These authors contributed equally to this work.

## Abstract

This study aimed to estimate the genetic parameters and breeding values of milk yield traits of Holstein cows in Shandong Province using the best model identified by a comparison between a numbers of alternative random regression test day models (RRMs). The data included 585,702 test day records of milk yield in the first lactation of 88,215 Holstein cows, covering 219 cattle farms in Shandong Province during the period from 2005 to 2016. Different models were investigated, which differed in the number of knots of Spline functions to improve the fitting of population lactation curve and in orders (2, 3, or 4) of Legendre polynomials to fit additive genetic effect and permanent environmental effect. The optimal test day model was screened out by Akaike information criterion (AIC) and Bayesian information criterion (BIC) criteria. Detailed analysis of genetic parameters and accuracy of estimation of breeding values were performed using the optimal model. In the results, the optimal model (Sp15-La4-Lp3) for analyzing the milk yield data was the one with 15 knots of Splines, 4 orders of Legendre polynomials for additive genetic effect and 3 orders of Legendre polynomials for permanent environmental effect. Using the optimal model, estimates of additive genetic variances of milk yield at different days in milk (DIM) during the whole lactation ranged from 8.54 to 15.39, the permanent environmental variance ranged from 17.65 to 31.42. Correspondingly, the heritability ranged from 0.20 to 0.30, and repeatability ranged from 0.43 to 0.54. Rank correlations between EBV of bull with different number of daughters and the bull's parent average ranged from 0.79 to 0.94, and the correlations between EBV of bulls and the sire-maternal grandsire index ranged from 0.48 to 0.86. In conclusion, Sp15-La4-Lp3 could be the optimal model for estimation of genetic parameters and prediction of breeding values of milk in Shandong Holstein population. The amount of progeny information is critical to the conventional genetic evaluation of bulls. © 2021 Friends Science Publishers

**Keywords:** Genetic parameter; Holstein cattle; Milk yield; Random regression model

## Introduction

Data for test day milk of dairy cows were classically longitudinal or repeated measurements and there is a correlation between the test day records of the same animals (White *et al.* 1999). The test day model (TDM) was an alternative to the traditional genetic assessment model. The traditional genetic evaluation model required the conversion of multiple test day records during a lactation period to a standard milk yield of 305-d. The test day model could provide some advantages compared to traditional models. The test day model directly took each test day record into genetic evaluation, and does not need to convert it into lactation record, which reduced the error in calculation of the cumulative lactation record. In addition, the test day

model allowed to account for the difference of genes and gene effects on milk production in different lactation periods. Thus, the test day model could improve the accuracy of genetic evaluation. Several types of test day model had been proposed, such as multiple-trait reduced rank model (Meseret *et al.* 2015), repeatability model (Behzadi and Mehrpoor 2018), covariance function model (Kirkpatrick *et al.* 1994) and random regression model (Schaeffer 1994).

A random regression model (RRM) with a polynomial or other simple function was increasingly being used by animal breeders (Jamrozik and Schaeffer 1997; Meyer and Hill 1997). When the lactation curve was fitted by random regression model, the random effect part of the model usually adopted Legendre polynomial as covariates (Kirkpatrick *et al.* 1994). According to Meyer (2005),

changed in variance along a continuous scale, in general, could be well modeled by high-order Legendre polynomials, but these polynomials might have an over-fitting phenomenon at the beginning and end of the trajectory. Segmented Polynomials (Splines) could be used to replace higher order Legendre Polynomials. In Spline regression model, lactation curve was described through a series of specific points (knots). Each segment composed of node connections was part of a low-order polynomial. Cubic smooth Splines could be inlaid in hybrid model frames (Lin *et al.* 2020). Cubic Spline function was a smooth curve formed by piecewise cubic interpolation multinomial. The first and second derivatives on the curve were continuous (Wan *et al.* 2018).

In the study of RRM in dairy cattle, Legendre Polynomials and Splines were rarely used to estimate additive genetic effects and permanent environmental effects. Silvestre *et al.* (2005) estimated the genetic parameters of milk, fat and protein yields of Portuguese cows by using a test day model with uniform Spline of 12 knots. Pereira *et al.* (2012) compared RRM with Legendre polynomial to that with linear Spline function based on milk yield records of purebred dairy Gyr (*Bos indicus*) and hybrid cows (dairy Gyr × Holstein), and found that the linear Spline function model with 6 knots has the best fitting to the data. Pool and Meuwissen (1999) compared different TDM for their ability to fit lactation curve of milk yield, the result showed that the TDM with Legendre polynomial of order 5 was most suitable for the evaluation of breeding value. Bohmanova *et al.* (2008) compared four RRM based on either Legendre polynomials of order 4 or linear Splines of 4, 5, or 6 knots, and concluded that those RRM tended to be overestimate the additive genetic variance at both ends of the fitting trajectory. However, in the model inlaid with Spline, the overestimation was smaller, and the Spline with 6 knots had the best fitting to the data based on the model comparison criterion.

Lots of milk records have been collected in Chinese dairy herds, so genetic parameters should be updated regularly and Genetic evaluation should be performed timely. Moreover, it has not been found investigating the impact of Spline function on genetic evaluation in Chinese Holstein population. This study was in order to compare RRM with different knots of cubic Spline function and different orders of Legendre polynomial for estimation of genetic parameters and prediction of breeding value in Chinese Holstein population, using test day records of milk yield.

## Materials and Methods

### Data

Data were obtained from Dairy Cattle Research Centre DHI Lab, Shandong Academy of Agricultural Sciences. The data included 585, 702 test day records of milk yield in the first lactation of 88, 215 Holstein cows, covering 219 cattle

farms in Shandong Province during the period from 2005 to 2016. Data quality control was performed using criteria: DIM between 5 and 305, calving ages from 22 to 38 months, daily milk yield from 5 to 80 kg and cows with at least 3 records. The pedigree was traced back 3 generations, resulting in a pedigree file containing 193,156 animals. The general statistics of the data were shown in Table 1.

### Models

The data were analyzed using RRM embedded with Legendre polynomials and cubic Spline functions. The fixed effects of the RRM include Herd-Year-Season (HYS), days in milk and calving age. Random effects include cubic Spline knots effect, additive genetic effect and permanent environmental effect. The calving age was divided into four groups: age ≤ 24 months, 24 < age ≤ 28 months, 28 < age ≤ 32 months, and age ≥ 32 months. Residual variances in different lactation period were assumed to be heterogeneous with 10 classes of residual variances (5–34, 35–64, 65–94, 95–124, 125–154, 155–184, 185–214, 215–244, 245–274, and 275–305 DIM).

The model can be written as

$$y_{ijkl} = HYS_i + DIM_j + Age_k + SPL_x + \sum_{m=0}^{n_1} a_m X_m + \sum_{q=0}^{n_2} p_q X_q + e_{ijkl}$$

Where  $y$  was the test day record ;  $HYS$  was the effect of Herd-Year-Season;  $DIM$  was the effect of days in milk;  $Age$  was the effect of calving age,  $SPL$  was the random regression coefficient on cubic Splines to improve the fitting of population lactation curve;  $a$  and  $p$  were the random regression coefficient of additive genetic effect and permanent environmental effect, respectively;  $X$  was Legendre Polynomial;  $m$  and  $q$  were order of Legendre polynomials of additive genetic effect and permanent environmental effect, respectively; and  $e$  was residual effect. In matrix notation:

$$y = Xb + Ts + Za + Wp + e$$

Where  $y$  was the vector of observation;  $b$  was the vector of fixed effect;  $a$  was the vector of the random regression coefficient of additive genetic effect.  $s$  was the vector of cubic Spline random regressions coefficient;  $P$  was the vector of random regression coefficient of permanent environmental effect;  $X$ ,  $T$ ,  $Z$ , and  $W$  were incidence matrices corresponding to fixed effect, additive genetic, and permanent environmental effect, respectively; and  $e$  was the vector of residual effects. The variance of  $a$ ,  $s$ ,  $P$  and  $e$  was:

$$\text{var} \begin{bmatrix} a \\ s \\ p \\ e \end{bmatrix} = \begin{bmatrix} G \otimes A & 0 & 0 & 0 \\ 0 & \sigma_s^2 \otimes I_s & 0 & 0 \\ 0 & 0 & P \otimes I_m & 0 \\ 0 & 0 & 0 & R \otimes I_n \end{bmatrix}$$

The additive genetic and permanent environmental effects were fitted by Legendre polynomials, and the order of Legendre polynomials is 2, 3 or 4. The random effects in the model also included cubic Spline functions of  $n_s$  knots that were evenly spaced throughout the lactation period. Firstly, the model combines the cubic Splines of 20 knots with Legendre polynomials of different orders for empirical selection of optimal order of Legendre polynomials, so the model was run 9 times for 9 order combinations (3 for additive genetics  $\times$  3 for permanent environment) at the Splines of 20 knots. The best order combination of these nine models was selected using AIC and BIC criterion. Then the data were analyzed using the models with the best order combination together with different (4, 5, 6, 7, 10, 15, 25, 30) knots for Spline functions. These models were evaluated using AIC and BIC criterion to obtain the model with best combination of number of knots for Spline functions and number of orders for Legendre polynomials.

### Methods of model comparison

As mentioned above, the models with different knots of Splines and different orders of Legendre polynomial were evaluated using AIC and BIC.

Akaike information criterion (AIC) (Akaike 1974):

$$AIC = -2\log(L) + 2p$$

Where L is the maximum of the restricted likelihood function; P is the number of parameters in the model.

Bayesian Information Criterion (BIC) (Schwarz 1978):

$$BIC = -2\log(L) + p * \log(n)$$

Where L is the maximum of the restricted likelihood function; P is the number of parameters in the model; n is the degree of freedom of the residuals.

When the model converges, lower AIC and BIC are preferred.

### Estimation of variance components

Estimation of additive genetic and permanent environmental variance:

$$\sigma_a^2(d) = \phi_{(d)} G \phi'_{(d)}$$

$$\sigma_{pe}^2(d) = \phi_{(d)} P \phi'_{(d)}$$

Where G is the covariance (COV) matrix between the random regression coefficients of additive genetic effects; P is the COV matrix between the random regression coefficients of permanent environmental effects;  $\phi$  is the

coefficient matrix of the Legendre polynomial. Different models were carried out for the estimation of covariance components and prediction of breeding values by the DMU package. Estimation of heritability:

$$h_i^2 = \frac{\sigma_{a_i}^2}{\sigma_{pe_i}^2 + \sigma_{a_i}^2 + \sigma_{e_i}^2}$$

Where  $h_i^2$  is the heritability of the  $i$ -th test day,  $\sigma_{a_i}^2$  is the additive genetic variance of the  $i$ -th test day, and  $\sigma_{pe_i}^2$  is the permanent environmental variance of the  $i$ -th test day.  $\sigma_{e_i}^2$  is the residual variance of the  $i$ -th test day which is the residual variance of the residual group the test day belongs to, and the range of  $i$  is  $i = 5, 6, 7, \dots, 305$ . Estimation of repeatability:

$$PE_i^2 = \frac{\sigma_{pe_i}^2}{\sigma_{pe_i}^2 + \sigma_{a_i}^2 + \sigma_{e_i}^2}$$

Where  $PE_i^2$  is repeatability, that means ratio of permanent environment of variance to phenotypic variance at  $i$ -th test day,  $\sigma_{pe_i}^2$ ,  $\sigma_{a_i}^2$  and  $\sigma_{e_i}^2$  were same as above. Estimation of breeding value:

$$EBV = A\phi'$$

Where EBV is the matrix of EBV for each individual (row) at each test day (column), A is the random regression coefficient matrix of each order (column) and each individual (row) for additive genetic effect estimated by the above model and  $\phi'$  is the transposed coefficient matrix of the Legendre polynomials of each order (row) and each test day (column). The breeding value of lactation milk yield for each individual can be obtained by adding up the EBV of each test day.

### Assessment of the contribution of daughter information to accuracy of bull EBV

The importance of daughter records for accuracy of bull EBV was assessed with correlation (r) between bull's EBV of 305 d milk yield (EBV<sub>305</sub>) and the pedigree index. The correlation is equivalent to the ratio of accuracy of pedigree index to accuracy of bull EBV. The contribution of daughter records to accuracy can be measured as (1-r)/r. In this study, Spearman Rank Correlation (Zar 2004) was used.

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Where,  $r_s$  was rank correlation coefficient,  $d$  was the rank

difference of individual breeding value and pedigree index, and  $n$  was the number of individual.

**Results**

In this section, we present the results of model comparison. The abbreviations of different models had the form of Spx-Lay-Lpz, where Spx is the number of Spline knots, Lay is the Legendre polynomial orders to fit additive genetic effect, Lpz is Legendre polynomial orders to fit permanent environmental effect.

**Goodness of fit of models**

Results of the models with the same 20 knots of Spline and different Legendre polynomials orders were shown in Table 2, where *Sp20-LA3-LP4* and *Sp20-La4-Lp4* did not converge. Given the Spline with 20 knots and different orders of Legendre polynomials, *Sp20-La4-Lp3* had the smallest AIC and BIC values, compared to all other converged models, indicating that this model had a better fit to the data than other models. It suggested that Legendre polynomials of order 4 werer optimal for additive genetic effect, and the Legendre polynomial of order 3 was optimal for permanent environmental effect, given the Splines with 20 knots.

Table 3 present AIC and BIC for models with different numbers of Spline knots and the selected optimal Legendre polynomial orders, *La4-Lp3*. All the models in Table 3 converged, in which *Sp15-La4-Lp3* showed a minimum value of AIC and BIC, indicating that this model fitted data better than other models. In other words, 15 knots for Splines was the optimal, given *La4-Lp3*.

**Comparison of heritabilities estimated from different models**

Fig. 1 shows a large difference in heritability of test day milk yield between models with 20 Spline knots but different orders of Legendre polynomials. It was observed that the heritabilities estimated by all the models changed over DIM of lactation and appeared a curve with multiple inflections. The number of inflection points of the heritability curves increased with the order of Legendre polynomials, and thus the inflection points on DIM were different in the heritability curve among the different models. The curve amplitude of *SP20-La4-Lp3* was smaller than the other model curves, indicating more smooth change in estimated heritability over DIM. The heritabilities estimated using *Sp20-La4-Lp3* ranged from 0.20 to 0.30.

The estimated heritabilities from the models with the same order of Legendre polynomials (*La4-Lp3*) but different number of Spline knots were almost the same in each DIM. The small difference was observed only in the early lactation period. The estimated heritabilities from

**Table 1:** Descriptive statistics of the test day record data

Item	Milk
Min (kg)	5.00
Max (kg)	79.94
Number of Records	585702
Average Yield (kg)	24.37
Standard Deviation	8.28
Coefficient of Variation	0.34

**Table 2:** AIC, BIC statistics of the models with the same 20 knots of Splines and different orders of Legendre polynomials

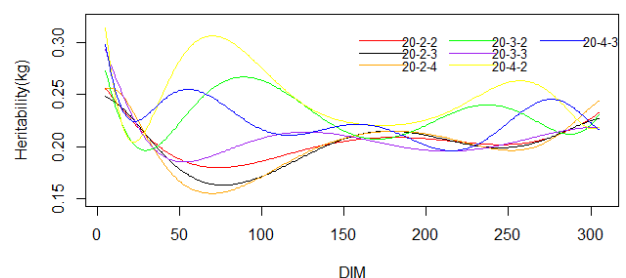
Model	$Spl^1$	$La$	$Lp$	$NP^2$	AIC	BIC	Converged
Sp20-La2-Lp2	20	2	2	23	2499477.24	2499736.58	YES
20-2-3	20	2	3	27	2494655.40	2494959.84	YES
20-2-4	20	2	4	32	2492389.73	2492750.55	YES
20-3-2	20	3	2	27	2494417.48	2494721.92	YES
20-3-3	20	3	3	31	2494023.97	2494373.51	YES
20-3-4	20	3	4	36	2491800.03	2492205.95	NO
20-4-2	20	4	2	32	2492099.31	2492460.12	YES
20-4-3	20	4	3	36	2491639.53	2492045.45	YES
20-4-4	20	4	4	41	2491325.28	2491787.58	NO

Define superscript 1 and 2

**Table 3:** AIC, BIC value of models with different knots of Splines but the same order of Legendre polynomials

Model	$P^1$	AIC	BIC	Converged
4-4-3	36	2492851.09	2493257.01	YES
5-4-3	36	2492656.54	2493062.46	YES
6-4-3	36	2492388.14	2492794.06	YES
7-4-3	36	2492208.63	2492614.55	YES
10-4-3	36	2491675.78	2492081.7	YES
15-4-3	36	2491632.29	2492038.21	YES
25-4-3	36	2491643.55	2492049.47	YES
30-4-3	36	2491642.64	2492048.56	YES

<sup>1</sup>P=number of parameters; AIC = Akaike information criterion; BIC=Bayesian Information Criterion; Converged=Model converge



**Fig. 1:** Heritabilities of milk yield at different DIMs, estimated using the models with 20 knots of Splines and different orders of Legendre polynomials

*Sp4-La4-Lp3* and *SP5-La4-LP3* ranged from 0.19 to 0.30 and those of the other models ranged from 0.20 to 0.30.

**Variance components and genetic correlations between DIMs estimated using the optimal model *Sp15-La4-Lp3***

The additive genetic variances, permanent environmental variances, phenotypic variances of test day milk yield estimated by the optimal model are shown in Table 4.

**Table 4:** Additive genetic variance, permanent environmental variance, phenotypic variance, heritability, and the ratio of permanent variance to phenotypic variance at different DIMs, estimated from the optimal model (Sp15-La4-Lp3) optimal model

DIM <sup>1</sup>	Milk				
	$\sigma_a^2$	$\sigma_{PE}^2$	$\sigma_p^2$	$h^2$	PE <sup>2</sup>
5	15.39	22.48	51.67	0.30	0.44
35	9.41	17.73	40.94	0.23	0.43
65	11.04	18.63	40.94	0.25	0.43
95	9.74	19.67	43.21	0.23	0.46
125	8.90	19.42	42.12	0.21	0.46
155	9.28	19.02	42.10	0.22	0.45
185	9.11	19.56	42.47	0.21	0.46
215	8.54	21.01	43.35	0.20	0.48
245	9.70	22.50	46.00	0.21	0.49
275	12.18	23.97	49.96	0.24	0.48
305	12.50	31.42	57.72	0.22	0.54

<sup>1</sup>DIM=Days in milk; AG=additive genetic variance; PE=permanent environmental variance; P=phenotypic variance;  $h^2$ =heritability; REP =repeatability

**Table 5:** Genetic correlation (below diagonal) and phenotypic correlation (above diagonal) between different DIM milk yields

	DIM	5	30	60	90	120	150	180	210	240	270	305
	5	...	0.55	0.34	0.25	0.25	0.21	0.20	0.20	0.15	0.10	0.06
	30	0.60	...	0.61	0.53	0.44	0.36	0.30	0.27	0.25	0.22	0.14
	60	0.16	0.87	...	0.65	0.65	0.47	0.39	0.34	0.31	0.28	0.20
	90	0.16	0.69	0.93	...	0.65	0.65	0.50	0.50	0.37	0.32	0.25
	120	0.01	0.69	0.71	0.91	...	0.65	0.59	0.52	0.44	0.37	0.30
Milk	150	0.01	0.24	0.47	0.74	0.95	...	0.65	0.60	0.52	0.43	0.35
	180	0.05	0.17	0.47	0.63	0.88	0.98	...	0.66	0.59	0.59	0.41
	210	0.02	0.23	0.47	0.62	0.80	0.88	0.95	...	0.66	0.61	0.49
	240	-0.03	0.32	0.49	0.62	0.64	0.67	0.76	0.92	...	0.69	0.57
	270	-0.06	0.34	0.49	0.48	0.47	0.47	0.57	0.79	0.96	...	0.67
	305	0.03	0.15	0.15	0.27	0.35	0.43	0.55	0.73	0.85	0.90	...

**Table 6:** Rank correlation of EBVs of bull with parent average and bSire-MGS index

Number of daughters	Number of bulls	Coefficient of Rank Correlation	
		Parent average	Sire-MGS index
1-10	899	0.94	0.86
11-20	179	0.88	0.74
21-30	105	0.86	0.68
31-50	145	0.88	0.70
51-100	147	0.87	0.69
101-200	80	0.79	0.48
201-	61	0.82	0.51

The additive genetic variance was higher at DIM 5–65 and DIM 275–305 than the DIMs in the middle lactation. The permanent environmental variance was higher at DIM 5(22.48) and DIM 305(31.42) than the DIMs at the middle lactation. The phenotypic variance was higher at DIM 5(51.67) and DIM 305(57.72) than the DIM at the middle lactation. The estimated heritability was the highest (0.30) at DIM5 and the lowest (0.20) at DIM 215. The estimated repeatibilities was the highest (0.54) at DIM305 and the lowest (0.43) at DIM35.

The genetic and phenotypic correlations among DIMs with an interval about 30 days are shown in Table 5, which ranged from -0.06 to 0.98 and from 0.06 to 0.69, respectively. The maximum value of genetic correlation coefficient was between DIM155 and DIM185, and the minimum value was between DIM5 and DIM270. The maximum of phenotypic correlation coefficient was between DIM240 and DIM270, and the minimum value was between DIM5 and DIM305. The genetic and phenotypic correlation of milk yield showed the same

pattern, *i.e.*, the correlation coefficient decreased with the increase of DIM interval. The smallest genetic correlation coefficient between two DIM with interval about 30 days was 0.60 at the beginning of lactation, and the highest was 0.98 in the middle lactation. The phenotypic correlation coefficients between two DIM with interval about 30 days were not big different in the whole lactation period, ranging from 0.55 to 0.67.

### Comparison of bull EBV accuracy and the pedigree index accuracy

The correlation between EBV<sub>305d</sub> of the bulls and the pedigree index is a measure of the ratio of pedigree index accuracy to bull EBV accuracy. As shown in Table 6, the rank correlations between EBVs of the bulls with different daughters and the parent average ranged from 0.79 to 0.94, indicating that the gains in accuracy from daughter information ranged  $\frac{1-0.94}{0.94} = 6.4\%$  from to  $\frac{1-0.79}{0.79} = 26.6\%$ ,

compared with parent average. The correlations between bull EBV and the pedigree index calculated from sire and maternal grand-sire (Sire-MGS index) ranged from 0.48 to 0.86, indicating that the gains in accuracy from daughter information ranged from 16.3 to 108.3%, compared with Sire-MGS index. The rank correlation coefficient between EBVs of bulls and the bulls' parent averages was 0.94 for the bulls with less than 10 daughters, and the correlation decreased to 0.82 when the number of daughters increased to more than 200. The rank correlation coefficient between EBVs of the bulls and the Sire-MGS index was 0.86 for the bulls with less than 10 daughters, and the correlation decreased to 0.51 for the bulls with more than 200 daughters.

## Discussion

To improve breeding quality, optimize breeding and increase milk yield, many factors, such as lactation curve (Boga *et al.* 2020; Kul 2020), exogenous hormones (Murtaza *et al.* 2020), feeding management (Mobashar *et al.* 2018; Atasever *et al.* 2020; Khan *et al.* 2020), body conditions and characters (Kul *et al.* 2020), genetics and diseases (Erdem and Okuyucu 2020; Kuropatwinska *et al.* 2020) have been studied.

This study analyzed test day models using RRM model with different knots of Splines and different orders of Legendre polynomial. The results showed that Sp15-La4-Lp3 could be the optimal model for estimation of genetic parameters and prediction of breeding values of milk yield in Shandong Holstein population, based on AIC and BIC statistics.

Additive genetic, permanent environment and phenotypic variances showed concave curves with and moderate oscillations throughout lactation period. This trend is consistent with Silvestre *et al.* (2005) and Ren *et al.* (2017). The estimated heritabilities of the test day milk yield over the entire lactation period ranged from 0.20 to 0.30, which was fallen between the estimates reported by White *et al.* (1999), but higher than the estimates by Zaabza *et al.* (2018) and Pelmus *et al.* (2016).

As expected, the genetic and phenotypic correlation between milk yields at different DIMs decreased with the increase of DIM interval, which is consistent with the results of Fazel *et al.* (2017) and Wang *et al.* (2017). The genetic and phenotypic correlation coefficients between DIMs reported by Fazel *et al.* (2017) ranged from -0.035 to 0.98 and from 0.03 to 0.67, respectively, which were very similar to the results obtained in this study. Wang (2011) studied the genetic parameters of Sanhe cattle by random regression model and the results showed that the range of genetic and phenotypic correlation of milk yield traits between DIMs in the first lactation period was -0.50~0.94 and -0.13~0.73. However, the phenotypic correlations of milk yield between DIMs reported by White *et al.* (1999); Silvestre *et al.* (2005) were 0.40~0.75 and 0.32~0.78, respectively.

The results are somewhat different in different studies, possibly due to the differences in environment, management and genetic background of the populations as well as statistical model used.

Correlation between EBVs of the bulls and the pedigree index reflected the contribution of daughter information to bull EBV, the lower correlation, the larger contribution. As shown in this study, both correlation between Bull EBV and parent average and correlation between bull EBV and Sire-MGS index decreased with increasing number of daughters. The rank correlation between EBVs for 305 d milk yield predicted using 305 d milk yield records using Lactation Model and EBVs for 305 milk yield estimated using test day RRM for bulls increased from 0.86 to 0.95 (Padilha *et al.* 2016). This indicates that the larger number of daughters of the bull the greater the rank correlation coefficient of the estimated breeding values between different models. The results are consistent with those obtained in this study.

## Conclusion

In all models of this study, the order of Legendre polynomials had larger effect on goodness of fit and the estimates of genetic parameters than the number of Spline knots. Sp15-La4-Lp3 could be the optimal model for estimation of genetic parameters and prediction of breeding values of milk yield in Shandong Holstein population. With the increase of daughters of the bull, the correlation coefficient between EBVs of bulls and the pedigree index decreased and thus the accuracy of Bull EBV in relation to pedigree index increased, indicating the amount of progeny information is critical to the genetic evaluation of bulls.

## Acknowledgement

This work was supported by Natural Science Foundation of Shandong Province (ZR2016CM37), Key Research and Development Plan of Shandong Province (2018GNC113003), Agricultural improved breed project of Shandong Province (2016LZGC027), China Agriculture Research System (CARS-36), Breeding and demonstration of high yield and  $\beta$ -casein A2 dairy cows (2019B10018), Agricultural scientific and technological innovation project of Shandong Academy of Agricultural Sciences (CXGC2016A04).

## Author Contributions

Jianbin Li designed the experiment and guided the writing of the article, Wan Lv analyzed the data, Wan Lv and Gaozhan Cai wrote this article, Gaozhan Cai revised this article, Yujiao Wang, Xianguo Gu, Xiuxin Zhao, Xiaotao Liu, Rongling Li, Minghai Hou analyzed and discussed results.

## References

- Akaike H (1974). A new look at the statistical model identification. *IEEE T Automat Cont* 19:716–723
- Atasever S, AV Garipoglu, H Erdem (2020). Effects of feeding management on milk yield and composition traits in crossbred jersey cows. *Pak J Zool* 52:1511–1516
- Behzadi MRB, Z Mehrpoor (2018). Estimation of genetic parameters of first lactation test-day milk yield using random regression models in Iranian holstein cow. *J Anim Plant Sci* 28:24–32
- Boga M, KK Cevik, A Burgut (2020). Classifying milk yield using deep neural network. *Pak J Zool* 52:1319–1325
- Bohmanova J, F Miglior, J Jamrozik, I Misztal, PG Sullivan (2008). Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. *J Dairy Sci* 91:3688–3699
- Erdem H, IC Okuyucu (2020). Non-genetic factors affecting some colostrum quality traits in holstein cattle. *Pak J Zool* 52:557–564
- Fazel Y, MA Fozi, A Esmailzadeh, F Fazel, AM Niazi, S Rahmati, MI Qasimi (2017). Use of random regression test-day model to estimate genetic parameters of milk yield in Holstein cows. *Open J Anim Sci* 08:27–38
- Jamrozik J, LR Schaeffer (1997). Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *J Dairy Sci* 80:762–770
- Khan MZU, TN Pasha, K Javed, MJ Basra, MA Jabbar (2020). Effect of varying dietary levels of calcium and phosphorus on apparent digestibility and retention of macro and micro-minerals in lactating sahiwal cows. *J Anim Plant Sci* 30:527–536
- Kirkpatrick M, WG Hill, R Thompson (1994). Estimating the covariance structure of traits during growth and ageing illustrated with lactation in dairy cattle. *Genet Res* 64:57–69
- Kul E (2020). Determination of the best lactation curve model and lactation curve parameters using different nonlinear models for anatolian buffaloes. *Pak J Zool* 52:1291–1297
- Kul E, A Sahin, E Ugurlutepe, M Soydaner (2020). Association of change in body condition score with milk yield and reproduction traits of holstein cows. *J Anim Plant Sci* 30: 305–311
- Kuropatwinska I, MBogucki, M Miller, ASawa (2020). Relationships between milk electrical conductivity, daily milk yield, and milking ability of primiparous and multiparous cows. *J Anim Plant Sci* 30:298–304
- Lin F, Y Tang, Z Zhu, SP Azen (2020). Weighted quantile regression in varying-coefficient model with longitudinal data. *Comput Stat Data Anal* 145:106915–106926
- Meseret S, B Tamir, G Gebreyohannes, M Lidauer, E Negussie (2015). Genetic analysis of milk yield in first-lactation holstein friesian in ethiopia: a lactation average vs. random regression test-day model analysis. *Asian Aust J Anim Sci* 28:1226–1234
- Meyer K (2005). Random regression analyses using B-Splines to model growth of Australian Angus cattle. *Genet Sel Evol* 37:473–500
- Meyer K, WG Hill (1997). Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. *Livest Prod Sci* 47:185–200
- Mobashar M, M Tahir, S Javaid, MI Anjum, I Gul, N Ahmad, A Sami (2018). Nutritional evaluation of various stages of maturity of Oat Hay and its effect on milk production and composition in lactating Holstein Friesian Cows. *Pak J Zool* 50:2209–2216
- Murtaza S, A Sattar, N Ahmad, M Ijaz, M Akhtar, M Shahzad (2020). Consequence of exogenous administration of oxytocin on reproductive and productive parameters during postpartum involution period in newly calved Nili-Ravi Buffaloes. *Pak J Zool* 52:1239–1254
- Padilha AH, JA Cobuci, CN Costa, JB Neto (2016). Random regression models are suitable to substitute the traditional 305-day lactation model in genetic evaluations of Holstein cattle in Brazil. *Asian Aust J Anim Sci* 29:759–767
- Pelmus RS, MC Rotar, H Grosu, E Ghita, C Lazar (2016). Using spline test-day model for estimating the genetic parameters for cows milk yield. *Sci Papers Anim Sci Biotechnol* 49:69–73
- Pereira RJ, AB Bignardi, LE Faro, RS Vermeque, AEV Filho, LG Albuquerque (2012). Random regression models using Legendre polynomials or linear Splines for test-day milk yield of dairy Gyr (*Bos indicus*) cattle. *J Dairy Sci* 95:565–574
- Pool MH, THE Meuwissen (1999). Prediction of daily milk yields from a limited number of test days using test day models. *J Dairy Sci* 82:1555–1564
- Ren XL, AX Liu, X Li, X Zhang, YC Wang, HF Shao, CH Qin, Y Wang, W Wen, SL Zhang (2017). Genetic Parameters Estimation of test day milk yield in Holstein Heifers in Ningxia using a random regression test-day model. *Sci Agric Sin* 50:1885–1892
- Schaeffer LR (1994). Random regressions in animal models for test-day production in dairy cattle. *Proc 5<sup>th</sup> World Congr Genet Appl Livest Prod* 18:443–446
- Schwarz G (1978). Estimating the dimension of a model. *Ann Stat* 6:31–38
- Silvestre AM, F Petim-Batista, J Colaço (2005). Genetic parameter estimates of Portuguese dairy cows for milk, fat, and protein using a Spline test-day model. *J Dairy Sci* 88:1225–1230
- Wan L, L WenJiao, X Jie, Y Hai-Yu, Z Ru-Mei, B Peng, L Jianbi, NZ Jifeng (2018). Application of spline function in genetic evaluation of dairy cattle. *Chin Dairy Cattle* 2018:23–26
- Wang QS, HY Tang, GL Liu, YC Pan (2017). Comparison of heritability and genetic correlation for milk yields of first and second-lactation Shanghai Holstein cows. *J Shanghai Jiaotong Univ (Agric Sci)* 35:90–94
- Wang RJ (2011). Study of Genetic Analysis of Lactation Curve and Estimated of Genetic Parameters in Sanhe Cattle. *Ph.D. Thesis*. Inner Mongolia Agricultural University, Inner Mongolia, China
- White IM, R Thompson, S Brotherstone (1999). Genetic and environmental smoothing of lactation curves with cubic Splines. *J Dairy Sci* 82:632–638
- Zaabza HB, AB Gara, B Rekik (2018). Genetic analysis of milk production traits of Tunisian Holsteins using random regression test-day model with Legendre polynomials. *Asian Aust J Anim Sci* 31:636–642
- Zar JH (2004). Spearman Rank Correlation. In: *Encyclopedia of Biostatistics*, Armitage P, T Colton (Eds.). John Wiley & Sons, Ltd., New Jersey, USA