

<sup>1</sup>Martin-Luther-University Halle-Wittenberg, Institute of Animal Breeding and Husbandry with Veterinary Clinic, Germany

<sup>2</sup>Res. Inst. Anim. Prod., Uhřetín, Czech Republic

HANA KREJČOVÁ<sup>1</sup>, NORBERT MIELENZ<sup>1</sup>, JOSEF PŘIBYL<sup>2</sup> and LUTZ SCHÜLER<sup>1</sup>

## **Comparison of Breeding Values for Daily Gains of Bulls Estimated with Multi-Trait and Random Regression Models\***

### **Abstract**

In this study, random regression models with Legendre polynomials of the 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> degree (RR2, RR3 and RR4) are compared with regard to the estimation of breeding values for the average daily gain of Czech Pied bulls (Simmental type). The data were prepared such that a multi-trait model (MTM) could be used as reference model. For each bull, 8 repeated records or fewer were available for the testing period from the 12<sup>th</sup> to the 420<sup>th</sup> day of life. For the modeling of the expected value structure, the fixed regression coefficients of the Legendre polynomials were subordinated hierarchically to the herd-year-season effects (HYS). For the comparison of the random regression models with the reference model, rank correlations between the estimated breeding values of various animal groups were calculated and a variety of top-lists were analyzed. In general, models RR3 and RR4 returned higher rank correlations with MTM in comparison to model RR2. Additionally, the number of common animals in the 1% and 10% top-lists showed that models RR3 and RR4 are to be preferred over RR2 when it comes to the estimation of breeding values.

Key Words: bulls, daily gain, random regression models, breeding values

### **Zusammenfassung**

**Titel der Arbeit: Vergleich von Zuchtwerten für die tägliche Zunahme von Bullen geschätzt mit Mehrmerkmals- und Random-Regression Modellen**

In dieser Studie werden Random-Regression Modelle mit Legendre-Polynomen 2., 3. und 4. Grades (RR2, RR3 und RR4) zur Schätzung von Zuchtwerten für die tägliche Zunahme von Fleckviehbullen miteinander verglichen. Die Daten wurden so aufbereitet, dass ein Mehrmerkmalsmodell (MTM) als Referenzmodell verwendet werden konnte. Pro Bulle lagen maximal 8 wiederholte Leistungen im Prüfzeitraum vom 12. bis 420. Lebenstag vor. Zur Modellierung der Erwartungswertstruktur wurden die fixen Regressionskoeffizienten der Legendre-Polynome den Herden\*Jahr\*Saison-Effekten hierarchisch untergeordnet. Zum Vergleich der Random-Regression Modelle mit dem Referenzmodell wurden Rangkorrelationen zwischen den geschätzten Zuchtwerten für unterschiedliche Tiergruppen berechnet und verschiedene Toplisten ausgewertet. Die Modelle RR3 und RR4 lieferten im Vergleich zu Modell RR2 generell höhere Rangkorrelationen mit Modell MTM. Auch die Anzahl gemeinsamer Tiere in 1% und 10% Toplisten zeigte, dass die Modelle RR3 und RR4 bei der Zuchtwertschätzung gegenüber Modell RR2 zu bevorzugen sind.

Schlüsselwörter: Fleischrindbullen, tägliche Zunahme, Random-Regression Modelle, Zuchtwerte

### **Introduction**

The estimation of breeding values is a main component of animal breeding. Various issues, such as selection processes or the optimization and improvement of breeding programs, are closely interlinked with it. Nowadays, the estimation of breeding values, or more precisely, the prediction of the additive-genetic effects, and the estimation of the variance components are carried out simultaneously with linear mixed models using the BLUP- and REML-method. If records of repeated observations of a certain trait (gathered on different control days) are available per animal, test-day models are increasingly used (SWALVE, 1995; HORSTICK and DISTL, 2002; SCHAEFFER,

\* Supported by the Ministry of Agriculture of the Czech Republic (Project No. 0002701401).

2004). The study at hand is concerned with the estimation of breeding values for the daily gain of Czech Pied bulls (Simmental type). Before their use for breeding purposes, the bulls were subjected to a test of their individual performance. During their stay in the rearing stations, the bulls were weighted in monthly intervals; on the test day, the bulls were at different stages in their growth. Estimations of breeding values of Czech Simmental bulls based on growth curves were presented by KREJČOVÁ et al. (2003) and PŘIBYL et al. (2004).

As already shown in a first report (KREJČOVÁ et al., 2007), so-called random regression models are quite suitable for the genetic analysis of the data. These models allow the utilization of all available performance records without any pre-adjustment with regard to systematic influential factors, including environmental effects on the test day, and enable the estimation of breeding values for each age of the animals in the test period (LEGARRA et al., 2004). The average daily gains of all bulls can be described quite well with the help of Legendre polynomials of the 2<sup>nd</sup> to the 4<sup>th</sup> degree. The introduction of subject-specific random regression coefficients allows the modeling of correlative relationships between repeated performances of a certain bull and facilitates the estimation of breeding values at any arbitrary point in time during the test period.

The average daily gains calculated from data on three successive body weights gathered during a period from day 12 to day 420 after birth were additionally split into eight non-overlapping sub-periods. In doing so, the simultaneous genetic analysis with a multi-trait model (MTM) was made possible. The well-introduced MTM was then used as reference model for random regression models which differed in their polynomial degree (MALOVRH, 2003). The comparison of the genetic parameters, such as heritability and genetic correlations, was already published (KREJČOVÁ et al., 2007). The focus of this study lies on the estimation of breeding values for average daily gains during the test period for both model variants and on the comparison of the estimated breeding values using rank correlations and top-lists.

### Material and Methods

For the analysis, records of the body weight of 6,420 Czech Pied bulls were available. The data were gathered when the bulls were between 12 and 420 days old and with monthly intervals on average. With the help of three successive recordings of the body weight, daily average gains were calculated. Since the data were to be evaluated with an 8-trait model as well, the test period was divided into 8 non-overlapping sub-periods (see Table 1). The average daily gain of a bull, calculated within an interval of 51 days, represents one trait. In order to obtain a unique classification, the record with the smallest distance from the middle of the time interval was chosen. Thus, every bull disposes of one to eight records at the most on daily gains in various stages of growth.

Table 1

Number of records (N), means ( $\bar{x}$ ) in g and standard deviations (SD) for the trait daily gain within 8 subperiods (Beschreibende Statistik für das Merkmal tägliche Zunahme)

$\Delta t_k$	[12,62]	[63,113]	[114,164]	[165,215]	[216,266]	[267,317]	[318,368]	[369,420]
$t_k$	37	88	139	190	241	292	343	394
N	978	3948	5236	6052	6095	5610	4593	1729
$\bar{x}$	743.7	930.1	1144.3	1242.8	1288.3	1249.5	1124.9	955.7
SD	196.2	213.9	249.1	218.5	206.3	214.7	229.0	247.0

No information on the relationships of the dams was available. The only case to be considered in the pedigree was the occurrence of a dam with more than one male calf. Hence, the majority of relationships included in the pedigree are sire relationships. In total, 253 sires are contained in the pedigree, with every bull having an average of 25 half-brothers. No adequate records on the performance of the sires of the base generation were available, and thus, their breeding values could only be estimated based on the yields of their sons. The pedigree contained a total of 7,125 animals.

For the estimation of breeding values, a multi-trait model and three different random regression models (RR models), which differed in the degree of the Legendre polynomials, were used. A detailed model description can be found in our first report (KREJČOVÁ et al., 2007). With the help of the MTM, breeding values were estimated for each of the eight sub-periods, whose arithmetic mean was then considered the breeding value for daily gains during the age period from day 12 to day 420.

In contrast, the RR models return the bulls' breeding values for every age within the period from day 12 to day 420. Let  $\alpha_j = (\alpha_{0j}, \dots, \alpha_{nj})'$  be the vector of the random regression coefficients of bull  $j$  within a certain RR model and with the Legendre polynomials of  $n$ -th degree as covariates. Let  $\phi(t^*) = (\phi_0, \phi_1(t^*), \dots, \phi_n(t^*))'$  denote the vector of the corresponding Legendre polynomials to the standardized age  $t^*$ . Furthermore, let  $t_{\min}$  and  $t_{\max}$  be the respective minimum and maximum age of bulls within the test period. It follows that the breeding values (BV) of animal  $j$  at age  $t$  are given by:

$$(1) \quad BV_j(t) = \sum_{i=0}^n \alpha_{ij} \cdot \phi_i(t^*) \quad \text{with} \quad t^* = \frac{2 \cdot (t - t_{\min})}{(t_{\max} - t_{\min})} - 1$$

As a next step, let  $t_k$  (with  $k=1, \dots, m$ ) represent the respective middle of each of the 8 sub-periods. The breeding value for the average daily gain of a certain bull would then be calculated with the following simple formula:

$$(2) \quad BV_j = \frac{1}{m} \cdot \sum_{k=1}^m BV_j(t_k)$$

The above definition allows the comparison of the breeding values from the MTM with those from the RR models. Furthermore, the impact of breeding value curves with strong variations at the beginning and end of the test period on the average aggregate breeding value is eliminated.

For the basic statistical analysis of the data, the statistics package SAS was employed. The estimation of variance components and breeding values was performed with the help of the REML and BLUP method respectively, and here the program VCE5 (KOVAC et al., 2002) was used. For the  $n$ -th polynomial degree, the program VCE5 returns predictions for  $(n+1)$  animal-specific regression coefficients provided the "solution" option is included. The calculation of breeding values based on formulas (1) and (2) was realized with the help of SAS using the matrix programming language IML. In order to make sure that the fixed effects in the output of VCE5 are not subjected to any program-internal transformation, VCE5 was executed without any scaling, i.e. with the option "scaley non", whenever starting values were used for the estimation of the variance components.

For the purpose of comparison, the estimated breeding values of all models were merged into one record per animal. The MTM served as reference model, and records were sorted by the breeding values estimated with this model. The comparison of the

breeding values was performed with the help of rank correlations based on all animals and based on those animals that disposed of at least one record. In addition, the breeding values for the best 50, 20, 10, 5, 2 and 1% among the animals were compared using rank correlations. Furthermore, top-lists for the 1% and 10% of animals with the best breeding values were generated for all models, and the number of animals appearing both in the top-lists of MTM and in the lists for the RR models was calculated.

### Results and Discussion

The changes in the expected values and variances of two-dimensional random variables with normal distribution when using a one-sided truncation selection for a variable can easily be detected under the assumption that the basic population is of infinite size (RASCH and HERRENDÖRFER, 1990). Let  $\alpha$  be the proportion of animals whose breeding values estimated with the help of MTM lie above the corresponding truncation point, and let  $\sigma_{bv(MTM)}$  be the standard deviation of the breeding values for all animals estimated with MTM. Based on the truncation formulas for a normal distribution, it follows for the expected values  $E(\cdot)$  of the truncated variables:

$$(3) \quad E(BV_{MTM}) = \mu_{MTM} + d_s \cdot \sigma_{bv(MTM)} \quad \text{and} \quad E(BV_{RRM}) = \mu_{RRM} + \rho \cdot d_s \cdot \sigma_{bv(RRM)}$$

Here,  $d_s$  denotes the standardized selection differential (selection intensity) to a given proportion  $\alpha$ , and  $\rho$  stands for the correlation between the breeding values from MTM and the RR models. Table 2 shows the means and standard deviations of the breeding values for all animals as well as for the percentage rates of animals with the highest breeding values. The results from Table 2 confirm the relations resulting from the truncation formulas. With a decreasing proportion  $\alpha$  of the selected animals, the means increase as expected, whereas the standard deviation of the breeding values is reduced.

Table 2

Means ( $\bar{x}$ ) and standard deviations (SD) of breeding values for the top animals (in %) and for all animals using MTM and RR models for the estimation. (Mittelwerte und Standardabweichungen der Zuchtwerte für die besten Tiere (in %) und für alle Tiere geschätzt mit MTM und RRM)

top %	1		2		5		10		20		50		all	
N	71	143	143	143	356	356	713	713	1425	1425	3563	3563	7125	7125
	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
<b>MTM</b>	81.9	9.4	73.9	10.5	62.7	11.7	53.6	12.5	42.9	14.2	25.0	18.0	1.35	30.1
<b>RR2</b>	67.7	12.9	61.7	12.9	50.9	13.8	43.7	13.9	35.4	14.5	21.0	16.8	1.36	26.4
<b>RR3</b>	77.1	12.8	70.0	12.9	58.0	13.9	49.5	14.1	39.8	15.0	23.3	17.9	1.32	28.8
<b>RR4</b>	81.0	12.2	74.1	12.2	62.2	13.6	53.3	14.3	42.7	15.8	25.1	19.0	1.05	31.3

With the help of (3), it can be proven that the expected and the estimated means for MTM match quite well. Further, it follows from (3) that the higher the value of the correlation  $\rho$ , the lower should the deviation of the estimated means of the RR models from the means of MTM be, provided the means and variances are of the same base level. A comparison of the means listed in Table 2 shows that the closest match for all percentage rates is achieved for MTM and RR4. The rank correlations between the breeding values estimated with MTM and those estimated with the RR models are

listed in Table 3; they include those for all animals and those for the best 1, 2, 5, 10, 20 and 50% of all animals.

Table 3

Rank correlations between the breeding values estimated using MTM and those estimated using models RR2, RR3 and RR4 for all animals and for the best  $\alpha$  % of all animals. (Rangkorrelationen zwischen den Zuchtwerten geschätzt mit Modell MTM und mit den Modellen RR2, RR3 und RR4 für alle Tiere und für die  $\alpha$  %-besten Tiere)

top (%)		1	2	5	10	20	50	100
# of animals		71	143	356	713	1425	3563	7125
MTM	RR2	0.625	0.623	0.701	0.677	0.733	0.826	0.938
	RR3	0.790	0.755	0.776	0.781	0.834	0.903	0.969
	RR4	0.757	0.742	0.803	0.806	0.859	0.919	0.975
RR2	RR3	0.868	0.882	0.898	0.875	0.893	0.926	0.972
	RR4	0.561	0.625	0.675	0.672	0.745	0.831	0.937
RR3	RR4	0.741	0.774	0.780	0.789	0.852	0.914	0.972

For all percentage rates, models RR3 and RR4 showed higher rank correlations with MTM than RR2. For the 356 animals with the best breeding values in MTM (i.e. for the best 5% of all animals), correlations of 0.701, 0.776 and 0.803 between MTM and RR2, RR3 and RR4 were estimated. For percentage rates of 5, 10, 20, 50 and 100%, model RR4 provided the highest values of correlation. Only for rates of 1 and 2% model RR3 achieved slightly higher values (0.790 and 0.755) when compared with model RR4. The rank correlations between RR2 and RR4 turned out to be lower than the correlations between RR models with adjacent polynomial degrees.

The tables 3 and 4 show equivalent patterns. However, Table 4 considers only those animals that dispose of at least one record. Due to this requirement, the number of animals is reduced from 7,125 to 6,420. As a result, calculations of the rank correlations for the best 1, 2 and 5% of all animals determined with MTM could only use 64, 128 and 321 records respectively.

Table 4

Rank correlations between the breeding values estimated using MTM and those using models RR2, RR3 and RR4 for animals with at least one record and for the best  $\alpha$  % of all animals. (Rangkorrelationen zwischen den Zuchtwerten geschätzt mit Modell MTM und mit den Modellen RR2, RR3 und RR4 für Tiere mit mindestens einer Leistung und für die  $\alpha$  %-besten Tiere)

top (%)		1	2	5	10	20	50	100
# of animals		64	128	321	642	1284	3210	6420
MTM	RR2	0.611	0.613	0.698	0.660	0.729	0.824	0.938
	RR3	0.751	0.752	0.772	0.770	0.835	0.901	0.969
	RR4	0.751	0.768	0.791	0.801	0.855	0.917	0.975
RR2	RR3	0.871	0.880	0.898	0.874	0.892	0.925	0.972
	RR4	0.557	0.644	0.662	0.651	0.736	0.831	0.937
RR3	RR4	0.742	0.794	0.775	0.782	0.850	0.913	0.972

The rank correlations between the four investigated models for all animals as well as for the animals with at least one record show very similar tendencies across all percentage rates. As Table 4 shows, model RR2 delivered the least significant rank correlations with MTM and model RR4 returned the most significant correlations (for rates larger than 1%).

As mentioned already in section Material and Methods, each bull disposes of up to 8 records at the most. Therefore, rank correlations with the estimated breeding values of those animals with at least 1, at least 2, and up to those with exactly 8 records were calculated for the 4 models (see Table 5). For example, 6,163, 5,862 and 4,879 animals dispose of at least 3, at least 4 and at least 5 records respectively.

Table 5

Rank correlations between the breeding values estimated using MTM and those using models RR2, RR3 and RR4 for a given minimum number ( $N_{\min}$ ) of records per animal. (Rangkorrelationen zwischen den Zuchtwerten geschätzt mit Modell MTM und mit den Modellen RR2, RR3 und RR4 bei gegebener Mindestanzahl ( $N_{\min}$ ) von Leistungen pro Tier)

$N_{\min}$		=8	≥7	≥6	≥5	≥4	≥3	≥2	≥1
# of animals		128	1284	3210	4879	5842	6163	6356	6420
MTM	RR2	0.953	0.933	0.930	0.936	0.938	0.939	0.939	0.938
	RR3	0.977	0.964	0.964	0.967	0.969	0.969	0.969	0.969
	RR4	0.974	0.971	0.972	0.974	0.975	0.975	0.975	0.975
RR2	RR3	0.980	0.974	0.971	0.972	0.973	0.973	0.972	0.972
	RR4	0.948	0.934	0.929	0.935	0.937	0.938	0.938	0.937
RR3	RR4	0.975	0.973	0.970	0.971	0.972	0.973	0.972	0.972

From Table 5 it becomes obvious that models RR3 and RR4 exhibit larger rank correlations with MTM than RR2. Independent from all restrictions with regard to the minimum number of records per animal, all rank correlations remain on a high level.

With an increasing number of records per animal, the accuracy of the estimation of the breeding values increases. As a consequence, the rank correlations between MTM and the RR models for the 128 animals with 8 repeated records lie between 0.95 and 0.98. In analogy to Tables 3 and 4, models with directly adjacent polynomial degrees return higher correlations than models whose polynomial degrees are further apart. For minimum numbers of 1 to 7, model RR4 yields the highest rank correlations to the reference model MTM.

Using the breeding values from the four analyzed models, top-lists for the best 1% and 10% of all animals were compiled with consideration first given to all animals and then given to only those animals with at least one record (see Table 6).

Table 6

Number (N) and proportion (in %) of common animals in the 1% and 10% top-lists of the MTM and RR models for all animals and for those animals with at least one record. (Anzahl (N) und Anteil (in %) der gemeinsamen Tiere in den 1% und 10% Toplisten der RR-Modelle und des MT-Modells bei Berücksichtigung aller Tiere und der Tiere mit Leistungen)

model	all animals				animals with records			
	top 1% list (=71)		top 10% list(=713)		top 1% list (=64)		top 10% list(=642)	
	N	%	N	%	N	%	N	%
RR2	46	65	557	78	42	66	500	78
RR3	53	75	588	83	48	75	532	83
RR4	52	73	605	85	48	75	548	85

The top-lists based on the breeding values estimated with MTM were compared to those estimated using the RR models. In addition, the number of animals appearing both in the top-list of model MTM and in the top-list of one of the RR models was determined.

Model RR2 showed the smallest number of common animals in both top-lists. For the 1% and 10% top-lists for all animals, only a 65% and 78% proportion was achieved. In contrast, the relative proportions of common animals in the 1% and 10% top-lists for all animals reached values of 75% and 83% for model RR3 and values of 73% and 85% for model RR4. For animals with at least one record, at total of 48 and 532 common animals could be found in the 1% and 10% top-lists respectively of MTM and model RR3. When comparing the 10% top-lists of models RR2, RR3 and RR4 with the top-list of MTM, respective totals of 500, 532 and 548 common animals were determined.

### Conclusions

In this study, the data were prepared in such a way that an evaluation with a multi-trait model was made possible. In order to do so, the test period was divided into eight non-overlapping sub-periods of almost identical length, and only one record on a bull's performance was accepted per sub-period.

From the estimated rank correlations between the models as well as from the calculated numbers of common animals in different top-lists, it follows that models RR3 and RR4 yield breeding values with a significantly higher degree of similarity to the values from MTM than those yielded by model RR2. If the well-analyzed multi-trait model is chosen as reference model, RR models of 3<sup>rd</sup> and 4<sup>th</sup> polynomial degree are to be recommended for the estimation of breeding values for the average daily gain. The use of RR models becomes a necessity when all available records per animal are to be considered instead of only those with the smallest temporal distance to the middle of the 8 established sub-periods.

The selection and adjustment of models for the description of repeated records per animal begins with the specification of the structure of expected values and the covariance of possible competing models, is followed by the comparison of estimated  $h^2$  and variance functions, and should include a verification process with the help of the comparison of rank correlations and top-lists as demonstrated in this study.

### References

- HORSTICK, A.; DISTL, O.:  
Estimation of genetic parameters for test day results of milk performance in East Friesian milk sheep using Bayesian methods for longitudinal data. Arch. Tierz., Dummerstorf **45** (2002) 1, 61-68
- KOVAC, M., GROENEVELD, E.; GARCIA-CORTEZ, L.A.:  
VCE-5, a package for the estimation of dispersion parameters. Proc. 7<sup>th</sup> WCGALP, Montpellier (2002).
- KREJČOVÁ, H.; PŘIBYL, J.; MISZTAL, I.:  
Breeding value for growth curve of performance tested dual-purpose bulls. 54<sup>th</sup> Ann. Meet. EAAP Rome, Italy 31 August – 3 September. Book of abstracts No. 9 (2003), 86
- KREJČOVÁ, H.; MIELENZ, N.; PŘIBYL, J.; SCHÜLER, L.:  
Estimation of genetic parameters for daily gains of bulls with multi-trait and random regression models. Arch. Tierz., Dummerstorf **50** (2007) 1, 37-46
- LEGARRA, A.; MISZTAL, I.; BERTRAND, J.K.:  
Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle. J. Anim. Sci. **82** (2004), 1564-1571
- MALOVRH, S.:  
Genetic evaluation using random regression models for longitudinal measurements of body weight in animals. Dis. Ljubljana (2003).
- PŘIBYL, J.; KREJČOVÁ, H.; PŘIBYLOVÁ, J.:  
Curve of weight gains in bulls. 5<sup>th</sup> Ann. Meet. EAAP, September 5.-8., Bled (2004)

RASCH, D.; HERRENDÖRFER, G.:

Populationsgenetik und Züchtungsmethodik (Handbuch). Deutscher Landwirtschaftsverlag Berlin (1990)

SCHAEFFER, L.R.:

Application of random regression models in animal breeding. *Livest. Prod. Sci.* **86** (2004), 35-45

SWALVE, H.H.:

Test day models in the analysis of dairy production data - a review. *Arch. Tierz., Dummerstorf* **38** (1995), 591-612

Received: 2006-06-23

Accepted: 2006-12-05

Corresponding Author

Dr. NORBERT MIELENZ

Institute of Animal Breeding and Husbandry with veterinary clinic

Martin-Luther University of Halle-Wittenberg

Adam-Kuckhoff-Straße 35

D-06108 Halle / Germany

E-Mail: [norbert.mielenz@landw.uni-halle.de](mailto:norbert.mielenz@landw.uni-halle.de)