

Modellwahl zur Schätzung von Laktationskurven beim Milchrind mit der SAS Prozedur Mixed

Holger Dressel
 Martin-Luther-Universität Halle-Wittenberg,
 Institut für Agrar- und Ernährungswissenschaften
 Arbeitsgruppe Biometrie und Agrarinformatik
 Ludwig-Wucherer-Str. 82-85
 06108 Halle (Saale)

holger.dressel@student.uni-halle.de joachim.spilke@landw.uni-halle.de

Zusammenfassung

Die Modellierung einer populations- und einzeltierbezogenen nichtlinearen Laktationskurve erfolgt mit Hilfe eines linearen gemischten Modells. Die Wahl des Modells für die festen Effekte und Kovariablen basiert auf einem zweistufigen Optimierungsprozess bei Nutzung analytischer Kriterien. Stufe eins beinhaltet die Optimierung der Erwartungswertstruktur. Im vorliegenden Anwendungsfall bestand die Aufgabe vor allem in der Wahl des Modells zur Beschreibung der Laktationskurve. Auf Stufe zwei, der Optimierung der Kovarianzstruktur, wird das in Stufe eins erarbeitete Basismodell, das außer den festen Effekten zunächst nur einen zufälligen Resteffekt mit diagonaler Struktur beinhaltet, durch Aufnahme weiterer zufälliger Effekte bei Beachtung unterschiedlicher Kovarianzstrukturen verfeinert. Es zeigt sich, dass die Beachtung laktationsnummernspezifischer fester Regressionskoeffizienten zur Beschreibung der Populationskurve und tierspezifischer zufälliger Regressionskoeffizienten zur Beschreibung des tierindividuellen Verlaufs bei Einbeziehung von Kovarianzen zwischen den Regressionskoeffizienten zur besten Anpassung führen.

Schlüsselwörter: Gemischte lineare Modelle, Modellwahl, Erwartungswertstruktur, Kovarianzstruktur, PROC MIXED, MIXED-Prozedur

1 Einleitung

Die Grundlage jeder genetisch bedingten Merkmalsänderung mit populationsgenetischen Methoden ist eine gezielte Selektion und Anpaarung der positiv selektierten Tiere zur Erzeugung von Nachkommen. Beim gegenwärtigen Entwicklungsstand der Tierzucht kommt der Nutzung phänotypischer Beobachtungen zur Schätzung der genetischen Merkmalswerte eine überragende Bedeutung zu. Da neben den eigentlich interessierenden genetischen Effekten weitere Störgrößen zu berücksichtigen sind, resultieren oft komplexe biometrische Modelle, deren den jeweiligen Daten entsprechende Struktur im Allgemeinen nicht bekannt ist. Daher kommt der Aufgabe der Modellwahl eine große Bedeutung zu. Diesem Problemkreis ist die vorliegende Arbeit zuzuordnen. Für das Beispiel wiederholt erfasster Milchmengen bei Milchkühen soll gezeigt werden, wie bei Nutzung der SAS Prozedur MIXED die Modellierung der Erwartungswert- und Kovarianzstruktur erfolgen kann. Derartige Daten sind bei Beachtung der Einflussgrößen Tier, Laktationsnummer, Laktationsdynamik und Testtag adäquat zu modellieren. Dabei

bedienen wir uns gemischter linearer Modelle, die über eine entsprechend große Flexibilität verfügen und die Abbildung nichtlinearer Laktationskurven ermöglichen, ohne dabei den linearen Ansatz verlassen zu müssen. In der vorliegenden Analyse werden zunächst verwandtschaftliche Beziehungen zwischen den Tieren nicht berücksichtigt. Die Arbeitsergebnisse haben sowohl Bedeutung für tierzüchterische Fragestellungen, die Auswertung von Fütterungs- und Haltungsversuchen mit Milchkühen als auch für die Unterstützung des Herdenmanagements durch die Bewertung betrieblicher Laktationskurven.

2 Material

Für die Arbeit standen einzeltierbezogen täglich erfasste Milchmengen eines Landwirtschaftsbetriebes in der östlichen Altmark zur Verfügung (Landwirtschaftsbetrieb der Landesanstalt für Landwirtschaft, Gartenbau und Forsten in Iden). Die Beobachtungen wurden an einer dreimal täglich gemolkenen Milchviehherde der Rasse „Deutsche Holstein“ erhoben. In die Analyse gingen 544.923 tägliche Milchleistungen von 797 Tieren, erfasst in einem Zeitraum von April 2003 bis Oktober 2008, ein. Der Auswertung lagen Daten bzw. Tiere mit den folgenden Bedingungen zugrunde (Laktationstag = LT):

- Tiere mit einer Laktationslänge > 200 LT und ≤ 450 LT
- Leistungen mit dem LT > 6 post partum
- Milchmengen (y) > 10 kg/LT und < 80 kg/LT

Zusätzliche Plausibilitätskontrollen:

- Milchmenge < 65 kg/Tag für den Laktationsabschnitt: $120. \text{ LT} < x \leq 200. \text{ LT}$
- Milchmenge < 55 kg/Tag für den Laktationsabschnitt: $200. \text{ LT} < x \leq 300. \text{ LT}$
- Milchmenge < 50 kg/Tag für den Laktationsabschnitt: $300. \text{ LT} < x \leq 450. \text{ LT}$

Die statistischen Maßzahlen der Milchmenge sind in Tabelle 1 zusammen gestellt. Wegen abnehmender Beobachtungszahlen in zunehmenden Laktationsnummern mussten Leistungen ab Laktationsnummer fünf zu einer neuen Laktationsnummer „fünf“ zusammengefasst werden.

Tabelle 1: Statistische Maßzahlen der täglichen Milchmenge (kg)

Laktationsnummer	Anzahl Beobachtungen	Anzahl Tiere	\bar{y}	\min_y	\max_y
1	146268	560	30.7	10.1	70.2
2	126475	472	34.9	10.1	78.9
3	104895	383	35.7	10.1	78.2
4	75635	281	36.4	10.1	77.1
„5“	91650	183	35.9	10.1	73.4

3 Methoden

Mit Hilfe eines mathematisch – statistischen Modells und der Wahl geeigneter Kovariablen ist es möglich, den nichtlinearen Verlauf einer Laktationskurve abzubilden, ohne

dass der lineare Modellansatz aufgegeben werden muss. In dieser Untersuchung werden verschiedene Ansätze zur Modellierung der Laktationskurve gegenübergestellt und deren Anpassung bei Nutzung analytischer Kriterien verglichen. Dazu wurden die Ansätze von Ali und Schaeffer [1], Guo und Swalve [3] und das Legendre Polynom sechsten Grades nach Kirkpatrick u. a. [5] und Mielenz u. a. [6] untersucht. In Tabelle 2 sind die Kovariablen der untersuchten Modellansätze zusammen gestellt.

Tabelle 2: Kovariablen der untersuchten Modelle (LT = Laktationstag)

Kovariable	Ali und Schaeffer (1987)	Guo und Swalve (1995)	Legendre Polynom 6. Grades
x_0	1	1	$(1/2)^{0,5}$
x_1	LT/LT_{\max}	LT	$(3/2)^{0,5} \cdot t$
x_2	$(LT/LT_{\max})^2$	$\sin(LT/100) \cdot LT^2$	$(5/2)^{0,5} \cdot (3/2 \cdot t^2 - 0,5)$
x_3	$\log(LT_{\max}/LT)$	$\sin(LT/100) \cdot LT^3$	$(7/2)^{0,5} \cdot 1/2 \cdot (5t^3 - 3t)$
x_4	$(\log(LT_{\max}/LT))^2$	$e^{-0,055LT}$	$(9/2)^{0,5} \cdot 1/8 \cdot (35t^4 - 30t^2 + 3)$
x_5			$(11/2)^{0,5} \cdot 1/8 \cdot (63t^5 - 70t^3 + 15t)$
x_6			$(13/2)^{0,5} \cdot 1/16 \cdot (231t^6 - 315t^4 + 105t^2 - 5)$ $t = 2 \cdot (LT - LT_{\min}) / (LT_{\max} - LT_{\min}) - 1$

Um eine Entscheidung treffen zu können, welche Kovariablen der drei genannten Modelle die beste Anpassung an die erhobenen Daten darstellen, wurde das durch Hurvich und Tsai (1989) modifizierte Akaike Informationskriterium verwendet. Dieser Entscheidungsvorgang basiert auf einer zweistufigen Vorgehensweise (vgl. Abschnitt 3.1/3.2).

3.1 Stufe 1: Optimierung der Erwartungswertstruktur

Der Modellierung der Erwartungswertstruktur liegt das nachfolgend dargestellte Basismodell zugrunde. Dabei ist y_{ijk} die Milchleistung von Tier k mit dem Laktationstag LT in der Laktation LAK i am Kontrolltag j . Diese Leistung wird als Realisation einer Zufallsvariable \underline{y}_{ijk} mit dem folgenden Modell angesehen:

$$(1) \quad y_{ijk} = \mu + KT_j + \sum_{r=0}^m b_{ir} \cdot x_r(LT) + \underline{e}_{ijk}^*$$

In diesem zunächst als Arbeitsmodell anzusehenden Modellansatz bezeichnen μ das allgemeine Mittel, KT_j den festen Effekt des Kontrolltages j , b_{i0} bis b_{im} die festen Regressionskoeffizienten von Laktationsnummer i und \underline{e}_{ijk}^* einen zunächst nicht weiter spezifizierten zufälligen Resteffekt. Weiter bezeichnen $x = (x_0, x_1, \dots, x_m)'$ einen Vektor von Kovariablen gemäß Tabelle 2 mit $m=5$ für die Modelle gemäß Ali und Schaeffer (1987) sowie Guo und Swalve (1995), $m=7$ für die Nutzung eines Legendre Polynoms 6. Grades (Kirkpatrick u. a. 1990, Mielenz u. a. 2007).

Bei Nutzung der SAS-Prozedur MIXED entspricht die Umsetzung von (1), beispielsweise für das Modell nach Ali und Schaeffer (1987):

```
PROC MIXED Method=ML;
  CLASS KT LAK;
  MODEL y = KT X0 (LAK) X1 (LAK) X2 (LAK) X3 (LAK) X4 (LAK);
```

In der MODEL - Anweisung sind X1, X2, X3 und X4 Kovariablen zur Beschreibung der laktationsnummernspezifischen Laktationskurve und KT der fixe Effekt des Kontrolltages.

Die Rangierung der Modelle erfolgte nach

$$AICC_{ML} = -2l_{ML} + \frac{2n(p_{rang} + q)}{n - p_{rang} + q - 1},$$

wobei l_{ML} den Wert der Likelihood der ML-Methode darstellt, p_{rang} der Rang der Versuchsplanmatrix für die festen Effekte, q die Anzahl der zu schätzenden Varianzkomponenten und n die Anzahl der Beobachtungen sind.

Außer dem Basismodell (1) und den in Tabelle 2 vorgestellten Kovariablen wurde ein Modell ohne die Berücksichtigung der Laktationsnummer formuliert und weitere Möglichkeiten der Kovariablenformulierung getestet.

3.2 Stufe 2: Optimierung der Kovarianzstruktur

Im Rahmen der weiteren Modellentwicklung ist die Abhängigkeit wiederholter Beobachtungen eines Tieres mit Hilfe der Kovarianzstruktur zu berücksichtigen. Dazu werden tierspezifische zufällige Regressionskoeffizienten als Abweichung von der in (1) modellierten laktationsspezifischen Populationskurve eingeführt. Sei $\underline{a}_{ik} = (a_{ik0}, \dots, a_{ikm})'$ ein laktationsspezifischer Vektor zufälliger Regressionskoeffizienten von Tier k , dann folgt in Erweiterung von (1):

$$(2) \quad y_{ijk} = \mu + KT_j + \sum_{r=0}^m b_{ir} \cdot x_r(LT) + \sum_{r=0}^m \underline{a}_{ikr} \cdot x_r(LT) + \underline{e}_{ijk}$$

mit $Var(\underline{a}_{ik}) = K_a$ und $Var(\underline{e}_{ijk}) = \sigma_e^2$

In Modell (2) werden alle zufälligen Effekte als normalverteilt angesehen. Weiterhin werden alle zufälligen Effekte verschiedener Tiere und zufälligen Effekte eines Tieres in verschiedenen Laktationen als unabhängig angesehen. Bei Verwendung von K_a und σ_e^2 können die phänotypische Varianz zu einem bestimmten Laktationsstadium bzw. die phänotypische Kovarianz innerhalb eines Tieres und Laktation zu verschiedenen Laktationsstadien ausgedrückt werden als:

$$Var(y(LT)) = x'(LT)K_a x(LT) + \sigma_e^2$$

$$\text{Cov}(y(LT_1), y(LT_2)) = x'(LT_1)K_a x(LT_2)$$

Bei Nutzung der SAS-Prozedur MIXED entspricht die Umsetzung von (2), beispielsweise für das Modell nach Ali und Schaeffer (1987):

```
PROC MIXED Method=REML;
  CLASS KT LAK TIER;
  MODEL y = KT X0(LAK) X1(LAK) X2(LAK) X3(LAK) X4(LAK);
  RANDOM INT X1 X2 X3 X4 / SUBJECT=TIER*LAK;
```

Während in der MODEL-Anweisung X1, X2, X3 und X4 Kovariablen wiederum zur Beschreibung der laktationsnummernspezifischen Laktationskurve sind und KT der feste Effekt des Kontrolltages ist, enthält die RANDOM-Anweisung den zufälligen Tiereffekt innerhalb einer Laktation und die dazugehörigen Kovariablen zur Schätzung tierspezifischer Verlaufskurven. Die Programmanweisung für (2) verwendet als Grundeinstellung die Varianzstruktur TYPE=VC. Das entspricht einer einfachen Struktur von K_a , in der alle Nichtdiagonalelemente Null gesetzt werden. Eine komplexere Struktur wird bei TYPE=UN geschätzt, indem auch Kovarianzen zwischen den zufälligen Regressionskoeffizienten zugelassen werden.

Die Rangierung der Modelle erfolgte nach

$$\text{AICC}_{\text{REML}} = -2l_{\text{REML}} + \frac{2qn^*}{n^* - q - 1},$$

wobei l_{REML} den Wert der Likelihood der REML-Methode repräsentiert, q die Anzahl der zu schätzenden Varianzkomponenten und n^* die Anzahl der Beobachtungen bzw. Tiere (bei subjektbezogener Modellformulierung) sind.

4 Ergebnisse und Schlussfolgerungen

Die Ergebnisse für AICC_{ML} der untersuchten Modelle sind in Tabelle 3 dargestellt. Es zeigt sich, dass das Legendre Polynom sechsten Grades einen Vorteil gegenüber den Modellansätzen von Guo und Swalve [3] und Ali und Schaeffer [1] besitzt. Der komplexere Ansatz des Polynoms muss also bezüglich der Anpassung der Kurve an die Beobachtungen besser geeignet sein, als die mit weniger Parametern arbeitenden erwähnten Ansätze. Es wurde davon abgesehen, Zwischenschritte der Modellentwicklung aufzuzeigen.

Tabelle 3: $AICC_{ML}$ der untersuchten Modellansätze (Erwartungswertstruktur)

Modelle	$AICC_{ML}$
Legendre Polynom 6. Grades	3637333
Guo und Swalve [2]	3637620
Ali und Schaeffer [1]	3637763

Zusätzlich zu den festen Effekten Testtag und Laktation und den zufälligen Resteffekten erfolgte im zweiten Modellierungsschritt die Aufnahme weiterer zufälliger Effekte wie Tier und Tier innerhalb einer Laktation. Aufgrund der günstigeren Konvergenz wurde bisher nur mit dem Modell von Ali und Schaeffer die Optimierung der Kovarianzstruktur durchgeführt. Die Ergebnisse der Integration weiterer zufälliger Effekte in das Modell von Ali und Schaeffer sind in Tabelle 4 aufgeführt. Es zeigte sich, dass die Einbeziehung aller Kovariablen und Varianzen unter Verwendung der Varianzstruktur TYPE=UN in der Prozedur MIXED zu den günstigsten $AICC_{REML}$ - Werten geführt hat.

Tabelle 4: $AICC_{REML}$ - Werte für die untersuchten Modelle (innerhalb Laktationsmodell nach Ali und Schaeffer [1])

Zufälliger Tiereffekt innerhalb Laktation mit den Kovariablen	Struktur von K_a	$AICC_{REML}$
Ohne	VC	3633922
X0	VC	3236443
X0, X1	VC	3101588
X0, X1, X2	VC	3047351
X0, X1, X2, X3	VC	3023885
X0, X1, X2, X3, X4	VC	3022013
X0, X1, X2, X3, X4	UN	2996351

Bei Nutzung der Prozedur IML und der Schätzungen für die festen Effekte und deren Kovarianzmatrix aus der Prozedur MIXED lassen sich die Verläufe der jeweiligen Laktationskurven und deren Konfidenzintervalle erzeugen. Die Ergebnisse sind beispielhaft für die erste und zweite Laktation in der Abbildung 1 und für die Laktationen drei, vier und fünf in Abbildung 2 dargestellt.

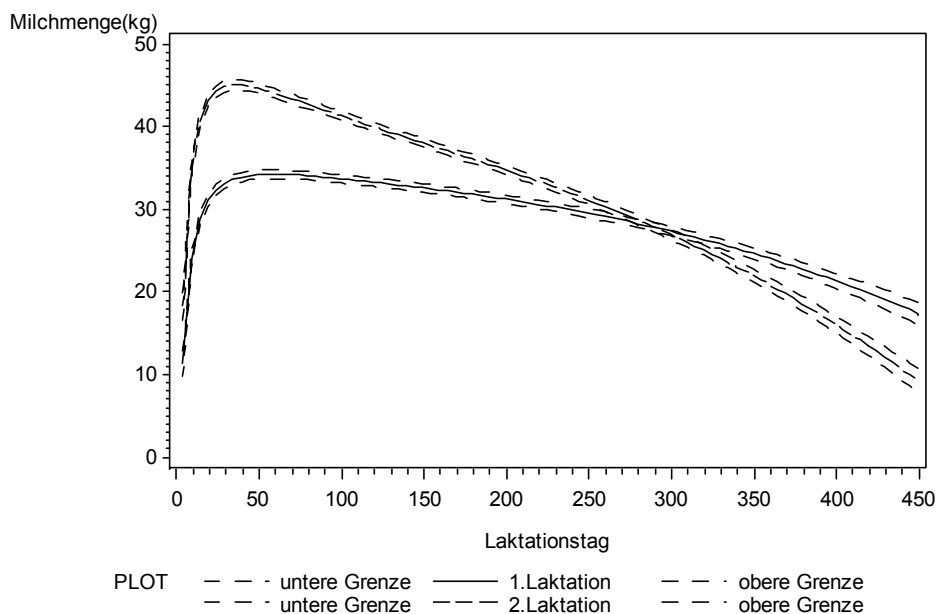


Abbildung 1: Laktationskurven für die tägliche Milchmenge der 1.-2. Laktation (Intervallgrenzen zweiseitig, $P=0,95$)

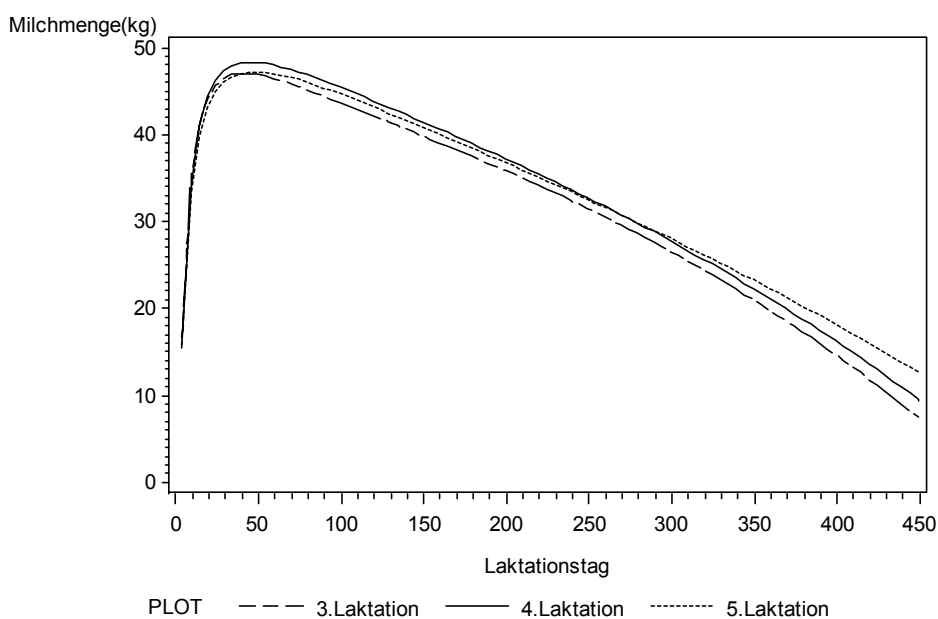


Abbildung 2: Laktationskurven für die tägliche Milchmenge der 3.-5. Laktation (aus Gründen der Übersichtlichkeit ohne Intervallgrenzen)

Es zeigt sich ein deutlicher Unterschied von der ersten Laktation zu den folgenden Laktationen bezüglich der Maxima der Kurven und der Persistenz. Dies drückt sich auch als signifikante Abweichung zwischen den Laktationen aus, hier in Abbildung 3 dargestellt am Beispiel der Differenz zwischen den Laktationen zwei und eins.

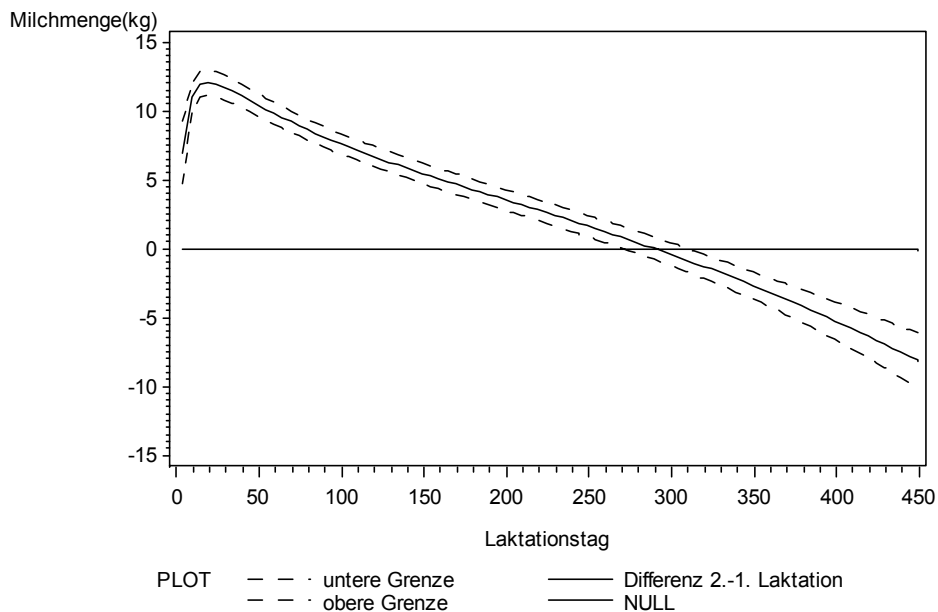


Abbildung 3: Differenz zwischen 2. und 1. Laktation für die tägliche Milchmenge (Intervallgrenzen zweiseitig, $P=0,95$)

Die Untersuchungen zeigen, dass bei Nutzung der verwendeten Modellansätze eine gute Abbildung von Laktationskurven bei Beachtung weiterer qualitativer Einflussfaktoren wie Kontrolltag, Laktationsnummer und Tier möglich ist. Das erarbeitete Modell kann vielfältig verwendet werden. Eine Nutzung ist beispielsweise für tierzüchterische Fragestellungen (jedoch bei zusätzlicher Einbeziehung der Verwandtschaftsmatrix!), Auswertung von Versuchen mit Milchkühen oder im Herdenmanagement eines Landwirtschaftsbetriebes möglich.

Literatur

- [1] Ali, T. E., Schaeffer, L. R. (1987): Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science* 67: 637-644.
- [2] Guo, Z. (1998): Modelle zur Beschreibung der Laktationskurve des Milchrindes und ihre Verwendung in Modellen zur Zuchtwertschätzung (Dissertation) Universität Göttingen.
- [3] Guo, Z., Swalve, H. H. (1995): Modelling of the lactation curve as a sub-model in the evaluation of test day records. *Proc. Interbull Meeting. Prague, Uppsala, Interbull Bull. No. 11.*
- [4] Hurvich, C. M., Tsai, C. L. (1989): Regression and time series model selection in small samples. *Biometrika* 76, 297-397.
- [5] Kirkpatrick, M., Lofsvold, D., Bulmer, M. (1990): Analysis of the Inheritance, Selection and Evolution of Growth Trajectories. *Genetics* 124: 979-993.

- [6] Krejcova, H., Mielenz, N., Pribyl, J., Schüler, L. (2007): Estimation of Genetic Parameters for Daily Gains of Bulls with Multi-Trait and Random Regression Model. Arch. Tierz. 50:1, 37-46
- [7] Spilke, J., Mielenz, N, Krause, S, Schüler, L. (2009): Statistical modeling for growth data in linear mixed models – Implications derived from an example of a population comparison of Golden Hamsters. Arch. Tierz. 52:1, 85-100