

Vergleich von Fixed- und Random-Regression Modellen bei verschiedenen Funktionsansätzen für Laktationskurven zur Vorhersage von Milchleistungen

Christian Ammon, Joachim Spilke
Arbeitsgruppe Biometrie und Agrarinformatik
Martin-Luther-Universität Halle-Wittenberg
Ludwig-Wucherer-Straße 82-85
D-06108 Halle
ammon@landw.uni-halle.de
spilke@landw.uni-halle.de

Abstract: Mixed linear models can be used to improve the informational value of milk yield forecasts. For this purpose two different functional approaches for modelling lactation curves are as well compared as three linear mixed models with varying random effects of individual animals and lactation numbers. It can be shown that more complex random regression models fit significantly better than fixed regression models.

1 Einleitung

Die Wirksamkeit der Nutzung beim Milchrind auf der Einzeltierebene zunehmend verfügbarer Daten ist vor allem daran gebunden wie es gelingt, diesen Datenbestand auch für einzeltierbezogene Entscheidungen zu verwenden. Ein großer Umfang verfügbarer Daten allein bringt keinen Vorteil, wenn Strategien zu seiner kontinuierlichen Nutzung für eine Entscheidungsunterstützung fehlen. Dabei kommt der Milchmenge eine überragende Bedeutung für die Bewertung des herden- und tierbezogenen Managements zu. Die erfassten Leistungen unterliegen komplexen Einflüssen (Testtag, Laktationsnummer und -dynamik, tierspezifische Effekte), ohne deren Berücksichtigung eine Verwendung der Daten zur Brunst- und Krankheitsdiagnose eingeschränkt ist. Ein Weg zur Verbesserung der Aussagefähigkeit verfügbarer Leistungen besteht in der Nutzung gemischter linearer Modelle zur Ermittlung von Leistungsvorhersagen. Zu diesem Zweck werden zum einen zwei Funktionsansätze zur Modellierung der Laktationskurven miteinander verglichen und zum anderen durch unterschiedliche Modellformulierungen verschiedene tier- und laktationsspezifische Effekte berücksichtigt.

2 Material und Methoden

Die verwendeten Daten wurden über einen Zeitraum von zwei Jahren (24.04.2002 bis 18.04.2004) von einem Milchviehbetrieb mit einem automatischen Melksystem und einer durchschnittlichen Herdengröße von 160 erfasst. Betrachtet wurde die auf 24-h-Intervalle umgerechnete Milchmenge von insgesamt 358 Tieren. Zur Verfügung standen über 69500 Datensätze. Es wurden vier Laktationsklassen gebildet: die Laktationen eins bis drei jeweils einzeln, während alle Datensätze der Laktationen vier und höher die vierte Gruppe bildeten. Diese Vorgehensweise war erforderlich, da Beobachtungen mit höheren Laktationsnummern nur noch mit geringer Häufigkeit vorlagen. Die Laktationskurven werden mit Covariablen modelliert, deren Regressionskoeffizienten (RK) als feste Effekte, zusätzlich auch als zufällige Effekte einbezogen werden (vgl. Tabelle 1). Der Ansatz nach Ali und Schaeffer sieht vier Covariablen vor ($X_1 = \text{Laktationstag}/400$; $X_2 = (X_1)^2$; $X_3 = \ln(400/\text{Laktationstag})$; $X_4 = (X_3)^2$) [AS87], nach Wilmlink werden zwei Covariablen zur Anpassung des Laktationsverlaufes genutzt ($X_1 = \text{Laktationstag}$; $X_2 = e^{-0.05 \text{ Laktationstag}}$) [Wi87]. Die Berechnung der Modelle wurde mit PROC MIXED aus dem Statistikpaket SAS vorgenommen. Ausgangspunkt ist ein gemischtes lineares Modell der Form

$$\underline{y} = X\underline{\beta} + Z\underline{u} + \underline{e}$$

mit \underline{y} als Vektor der Beobachtungen, X und Z als Designmatrizen der festen bzw. zufälligen Effekte, $\underline{\beta}$ und \underline{u} als Vektor der festen und zufälligen Effekte sowie \underline{e} als Vektor der zufälligen Resteffekte. Testtag und Laktationsnummer sowie die RK zur Anpassung der Laktationskurve innerhalb einer Laktationsgruppe bilden die festen Effekte. Je nach Auswahl und Gruppierung der zufälligen Effekte nimmt die Komplexität des Modells zu, d.h. es werden mehrere Kovarianzparameter geschätzt. Durch die zusätzliche Berücksichtigung zufälliger RK werden einzeltierspezifische Laktationskurven ermittelt.

Modell	1: Fixed Regression	2: Random Regression I	3: Random Regression II
Feste Effekte	Laktationsnummer		
	Testtag		
	RK innerhalb der Laktationsnummer		
Zufällige Effekte	Tier		Tier in Laktationsnummer
		RK zur Modellierung der Laktationskurve innerhalb Tier	
			RK zur Modellierung der Laktationskurve innerhalb Tier und Laktation
	Rest		

Tabelle 1: Feste und zufällige Effekte der untersuchten Modelle

In Modell 1 wird für die gesamte Herde eine Laktationskurve für jede Laktationsklasse angepasst. Für jedes Tier wird ein zufälliger Effekt geschätzt, um den die Laktationskurve tierspezifisch entsprechend verschoben wird. Modell 2 hat eine Schätzung einer Laktationskurve für jedes Tier zum Ergebnis, da neben den festen RK zur Beschreibung der Laktationskurve auf Herdenniveau zusätzlich tierspezifische RK berücksichtigt werden. Die Laktationskurven zweier Tiere in derselben Laktation unterscheiden sich demnach nicht durch eine über den Laktationsverlauf konstante Differenz der Tiereffekte wie in Modell 1, sondern können auch unterschiedliche Steigungen und Gefälle aufweisen. In Modell 3 wird auch für jede Laktation innerhalb eines Tieres zusätzlich ein separater zufälliger Effekt beachtet (Tabelle 1). Der Vergleich der Modelle erfolgt mit Hilfe von Kenngrößen für die Anpassungsgüte. Es wurde die -2 Log Likelihood und das Akaike-Information-Criterion (AIC) betrachtet [WK90]. Für beide Größen gilt, dass die Anpassung eines Modells umso besser ist, je niedriger der Wert des Kriteriums ausfällt. Um eine Modellbewertung bei Beachtung fester und zufälliger Effekte vornehmen zu können, wurde für die Aufstellung der Likelihood-Funktion die Methode ML (Maximum Likelihood) gegenüber der sonst üblichen Methode REML (Restricted Maximum Likelihood) zur Schätzung der Varianzkomponenten verwendet. Der berechnete Wert der Likelihood-Funktion kann zur Bewertung der Modellanpassung zwischen den Funktionen nach Ali und Schaeffer sowie nach Wilmink verwendet werden. Zusätzlich wurde innerhalb der Modelle mit gleichen Funktionsansätzen zur Laktationskurvenanpassung mit dem Likelihood-Quotienten-Test überprüft, ob sich die Modellstufen signifikant voneinander unterscheiden.

3 Ergebnis

Modell	Ali & Schaeffer			Wilmink		
	-2 Log Likelihood	AIC	Likelihood -Quotient	-2 Log Likelihood	AIC	Likelihood -Quotient
Modell 1: Fixed Regression	392197.7	393395.7	Vergleich mit ...	395306.3	396482.3	Vergleich mit ...
Modell 2: Random Regression I	367076.2	368282.2	Modell 1: 25121.5* (5 FG)	374553.5	375739.5	Modell 1: 20752.8* (3 FG)
Modell 3: Random Regression II	347744.1	348972.1	Modell 2: 19332.1* (15 FG)	352203.7	353407.7	Modell 2: 22349.8* (9 FG)

*signifikant bei $\alpha=5\%$

Tabelle 2: Anpassungsgüte der Modelle

Für beide Funktionsansätze gilt, dass die Anpassungsgüte von Modell 1 bis Modell 3 zunimmt. Über alle drei Modellstufen hinweg liefert der Funktionsansatz von Ali und Schaeffer jedoch die besseren Resultate im Vergleich zur Funktion nach Wilmlink. Beim Likelihood-Quotienten-Test wurde das jeweils komplexere Modell mit dem restriktiven Vorgänger verglichen (χ^2 -Test). Die Freiheitsgrade ergeben sich aus der Anzahl Restriktionen des restriktiven gegenüber dem komplexeren Modell. Im Unterschied zur $-2 \log$ Likelihood berücksichtigt AIC zusätzlich die Modellkomplexität als Strafterm ($AIC = -2 \log L + 2d$ mit $d =$ Summe der festen Modellparameter und der effektiven Anzahl der Varianzkomponenten im Modell).

4 Diskussion und Schlussfolgerungen

Gegenüber der alleinigen Betrachtung der tierindividuellen Abweichungen von einer Laktationskurve im Fixed Regression Modell, die den Herdendurchschnitt für die jeweils betrachtete Laktation darstellt, sind die Random Regression Modelle mit der Möglichkeit der Einbeziehung laktations- und einzeltierspezifischer Effekte vorzuziehen. Der Funktionsansatz von Ali und Schaeffer mit vier Covariablen zur Modellierung erweist sich gegenüber dem Ansatz mit nur zwei Covariablen von Wilmlink als vorteilhaft und passt sich dem tatsächlichen Laktationskurvenverlauf besser an. Weitere Modellverbesserungen sind zu erwarten, wenn die Anzahl der Laktationsklassen erhöht wird.

Zum jetzigen Zeitpunkt sind günstigenfalls nur von zwei Laktationen je Tier weitgehend vollständige Daten vorhanden, daher erscheint es möglich, dass bei einer erhöhten untersuchten Anzahl von Laktationen je Tier die Unterschiede innerhalb eines Tieres und innerhalb aufeinander folgender Laktationen besser schätzen lassen und sich so eine weitere Verbesserung der Modellanpassung ergibt.

Danksagung

Dieses Projekt wird durch die Deutsche Forschungsgemeinschaft (DFG) gefördert (Projektkennzeichen SP 613/4).

Literaturverzeichnis

- [AS87] Ali, T.E.; Schaeffer, L.R.: Accounting for covariance among test day milk yields in dairy cows. *Canadian Journal of Animal Science* 67, 1987; S. 637-644.
- [WK90] Wada, Y.; Kashiwagi, N.: Selecting statistical models with information statistics. *Journal of Dairy Science* 73, 1990; S. 3575-3582.
- [Wi87] Wilmlink, J.B.M.: Comparison of different methods of predicting 305-day milk yield using means calculated from within-herd lactation curves. *Livestock Production Science* 17, 1987; S. 1-8.