

Möglichkeiten des Monitorings von Milchkuhherden durch die automatische Erstellung gemischter linearer Modelle am Beispiel der Kontrolle von Gruppenwechseln

Veit Zoche-Golob, Christian Köber

Zukunftsforum „Veredlungsland Sachsen 2020“
Sächsischer Landeskontrollverband e. V.
August-Bebel-Straße 6
09577 Lichtenwalde
{veit.zoche;christian.koeber}@rizu.de

Abstract: Gemischte lineare Modelle eignen sich häufig gut zur Beschreibung von Zusammenhängen mit mehreren Einflussfaktoren. Ihr Einsatz zum Monitoring in der Milchkuhhaltung setzt allerdings voraus, dass sie ohne Prüfung jedes einzelnen Modells durch einen Menschen automatisch erstellt werden können. Die Analyse der Auswirkungen von Gruppenwechseln auf die Milchleistung von Kühen wird als Beispiel eines Versuchs der Automatisierung vorgestellt.

1 Einleitung

Im Rahmen des vom Freistaat Sachsen geförderten Projekts „Zukunftsforum Veredlungsland Sachsen 2020“ des Sächsischen Landeskontrollverbands e. V. (LKV) wird das Angebot "fitness monitoring" zum Controlling in der Milchproduktion zur „Daten- und Informationsplattform Tier“ (DIT) weiterentwickelt [Zo11], [SL12]. Mit diesem serverbasierten Programm können komplexe statistische Auswertungen wie die Erstellung gemischter linearer Modelle durchgeführt und den Nutzern die Ergebnisse in verständlicher Form präsentiert werden. Die Untersuchung der Auswirkungen von Gruppenwechseln auf die Milchleistung soll als Beispiel für die Möglichkeiten der DIT vorgestellt werden.

Um Milchkühe bei einer Fütterung mit totaler Mischration laufend entsprechend ihrer Leistung zu füttern, müssen sie während der Laktation wenigstens einmal von einer Fütterungsgruppe in eine andere umgestellt werden [SP04]. Die Kontrolle der Auswirkungen von Gruppenwechseln auf die Milchleistung ist nur schwer möglich. Entweder muss sich der Landwirt auf seine subjektive Wahrnehmung verlassen oder die Leistung einiger Kühe über ein paar Tage verfolgen. Die natürliche Entwicklung der Milchleistung im Laktationsverlauf kann er dabei nicht berücksichtigen.

2 Material und Methoden

Zur Erprobung des Auswertungsprozesses werden die betrieblichen Aufzeichnungen der Gruppenzuordnung der Kühne und ihrer in den Milchleistungsprüfungen (MLP) gemessenen Milchmengen von 01.08.2010 bis 31.07.2012 von 16 Herden der Pilotbetriebe des Projekts verwendet. Durchschnittlich bestehen sie aus 613 Kühen (Standardabweichung (SD): 363), die eine mittlere 305-Tage-Leistung von 9391 kg Milch (SD: 854,6) haben. Der gesamte Untersuchungszeitraum wird in Teilzeiträume zu zwölf Monaten, zu sechs Monaten und zu drei Monaten unterteilt. Für jeden Teilzeitraum wird für jede Herde ein Datensatz mit folgenden Spalten erstellt und untersucht: Datum der MLP, Laktationsindex (Laktation pro Kuh), Laktationstag, Milchmenge und Gruppenwechsel. Wenn eine Kuh bei einer MLP in einer anderen Gruppe stand als bei der vorangegangenen MLP, wird der Gruppenwechsel als „Ausgangsgruppe -> Zielgruppe“ definiert, andernfalls als „ohne“ (Referenzlevel).

2.1 Statistisches Modell

Damit bei der Analyse die Veränderung der Milchleistung im Laktationsverlauf berücksichtigt werden kann, wird das von Ali und Schaeffer empfohlene lineare gemischte Regressionsmodell [AS87] erweitert zu:

$$y_{idg}(t) = GW_g + DAT_d + \sum_{r=0}^4 (\beta_r + a_{ri}) \cdot x_r(t) + e_{idg}$$

Hierbei sind:

- $y_{idg}(t)$ die beobachtete Milchmenge in Laktation i zum Laktationstag t geprüft an Tag d nach Gruppenwechsel g
- GW_g die festen Effekte der Gruppenwechsel
- DAT_d die festen Effekte des Tags der MLP
- β_r die festen und a_{ri} die zufälligen Regressionskoeffizienten des Laktationskurvenmodells für Laktation i
- $x_0(t) = 1; x_1(t) = t / t_{max}; x_2(t) = x_1(t)^2; x_3(t) = \ln(t_{max} / t); x_4(t) = x_3(t)^2$
- e_{idg} der zufällige Rest in Laktation i an Tag d nach Gruppenwechsel g.

Die Differenz der Parameter von GW = „ohne“ und eines Gruppenwechsels liefert dann den erwarteten Einfluss dieses Gruppenwechsels auf die Milchleistung.

Zur Ermittlung der Modellparameter wird das Paket lme4 des Open Source Statistikprogramms R 2.14.2 [RD12] verwendet.

2.2 Automatisierung der Modellerstellung in einem R-Skript

Zuerst sind die Daten für jede Herde und jeden Untersuchungszeitraum so vorzubereiten, dass mit hoher Wahrscheinlichkeit ein sinnvoll interpretierbares Modell erstellt werden kann. Deshalb werden Gruppenwechsel, die sehr selten vorkommen, und Laktationen, zu denen nur Daten einer MLP vorliegen, ausgeschlossen. Außerdem wird geprüft, ob ü-

berhaupt mindestens ein Gruppenwechsel im Datensatz vorkommt und ob der Datensatz genügend Beobachtungen enthält, dass die Parameterschätzung theoretisch möglich ist. Wenn die Regression von lme4 ohne Warnungen oder Fehler durchgeführt wird, werden die geschätzten Laktationskurven auf Plausibilität überprüft (Abbildung 1). Sie müssen zu Laktationsbeginn steigen, dürfen über die untersuchte Laktationslänge keine negativen Werte enthalten und am Laktationsende nicht ansteigen. Falls keine plausible Kurve ermittelt werden konnte, wird die Regression mit eingeschränkten Datensätzen wiederholt. Dazu werden Daten von extrem hohen Laktationstagen ausgeschlossen oder nur die häufigsten Gruppenwechsel verwendet bzw. beide Einschränkungen kombiniert. Kann auch dann keine plausible Kurve geschätzt werden, wird keine Auswertung ausgegeben. Jeder Fehler und jede verwendete Einschränkung werden protokolliert.

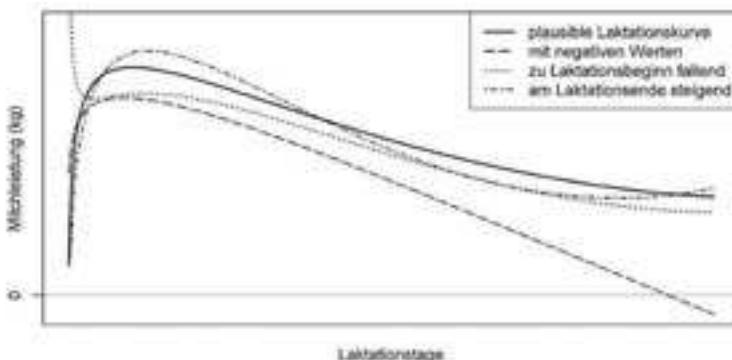


Abbildung 1: Geschätzte Laktationskurven

Die Ausgabe wird so vorbereitet, dass sie vom Anwender leicht interpretiert werden kann. In einer Tabelle werden alle untersuchten Gruppenwechsel mit der erwarteten Änderung der Milchleistung und dem mittleren Laktationstag der auf den Gruppenwechsel folgenden MLP mit Standardabweichung aufgelistet (Tabelle 1). Ist der Einfluss eines Gruppenwechsels geringer als ein eingestellter Schwellenwert oder nicht sicher genug von null verschieden, so wird er nicht ausgegeben.

Gruppenwechsel	Erwartete Änderung der Milchmenge	Mittlerer Laktationstag (SD)
02->01	-2,5	309 (83)
03->00	0,8	86 (22)
04->00	ohne deutlichen Einfluss	80 (39)

Tabelle 1: Auswertung der Gruppenwechsel (Beispiel der Ausgabe)

3 Ergebnisse und Diskussion

Datensätze über sechs (34,4%) oder zwölf Monate (43,8%) liefern deutlich häufiger zuverlässig interpretierbare Ergebnisse als solche über drei Monate (14,8%). Sechs Monate als Grundlage der Auswertung scheinen ein guter Kompromiss zwischen Aktualität

und Durchführbarkeit zu sein. Von den erfolgreich erstellten Auswertungen können 61,8% erst nach Ausschluss von Daten von extrem hohen Laktationstagen erstellt werden. Die anderen Einschränkungen der zu untersuchenden Daten tragen nicht zur Durchführbarkeit bei und könnten zur Beschleunigung des Prozesses übergegangen werden. Bei etwa der Hälfte der erfolglosen Versuche kann zwar die Regression durchgeführt werden, aber die geschätzten Kurven werden als „unplausibel“ verworfen. In diesen Fällen werden keine Aussagen über den Einfluss der Gruppenwechsel gemacht, damit sich die Anwender auf die dargestellten Auswertungen verlassen können. Die Ergebnisse zeigen, wie schwierig es ist, gemischte lineare Modelle automatisch zu erstellen. Eine große Herausforderung ist dabei, die Prüfung jedes Modells so zu automatisieren, dass möglichst keine unplausiblen Modelle akzeptiert und trotzdem kaum Modelle fälschlicherweise verworfen werden. Beim vorgestellten Verfahren werden noch sehr viele Modelle als „unplausibel“ eingestuft. Daher sollten die Prüfkriterien weiter verfeinert werden.

4 Ausblick: Einbindung des R-Skripts in die DIT

Durch die bestehende Infrastruktur des LKV können die erforderlichen Einzeltierdaten aus den lokal installierten Herdenmanagementprogrammen täglich in die Datenbank der DIT importiert werden [Be10]. Die Prozesse der DIT werden im .NET-Framework entwickelt. Mit der Bibliothek R.NET ist R in dieser Umgebung verfügbar [RN12]. Nachdem die Daten im festgelegten Format an R übergeben wurden, wird das beschriebene Skript aufgerufen. Dessen Rückgabe wird gespeichert und in der ASPX-Webapplikation ausgegeben. Obwohl noch nicht für jeden beliebigen Zeitraum eine Auswertung erstellt werden kann, erhalten Milchviehhalter so die Möglichkeit, als Anwender der DIT ihr Gruppenmanagement ohne zusätzlichen Aufwand zu kontrollieren.

Literatur

- [AS87] Ali, T.; Schaeffer, L.: Accounting for Covariances among Test Day Milk Yields in Dairy Cows. *Canadian Journal of Animal Science*, 67, 1987; S. 637-644.
- [Be10] Berger, H.: Praxisbericht zur Integration einer serviceorientierten Architektur für die Milcherzeugung beim Sächsischen Landeskontrollverband e.V. am Beispiel der Plattform fitness monitoring. In (Fähnrich, K.-P.; Franczyk, B. Hrsg.): INFORMATIK 2010: Service Science - Neue Perspektiven für die Informatik Band 1, Leipzig 2010. Köllen Verlag+Druck, Bonn, 2010; S. 173-178.
- [RD12] R Development Core Team: R: A Language and Environment for Statistical Computing. Wien, 2012. <http://www.R-project.org> (abgerufen am 06.11.2012).
- [RN12] R.NET. <http://rdotnet.codeplex.com> (abgerufen am 05.09.2012).
- [SL12] Sächsischer Landeskontrollverband e.V.: Veredlungsland Sachsen 2020. <http://www.veredlungsland.de/index.php?id=108> (abgerufen am 05.11.2012).
- [SP04] Spiekers, H.; Potthast, V.: Erfolgreiche Milchviehfütterung. DLG-Verlag, Frankfurt am Main, 2004.
- [Zo11] Zache, V.: fm-Milch – ein Monitoringinstrument für Milcherzeugerbetriebe. In (Fürll, M. Hrsg.): International Conference: Prophylaxis of herd and production diseases, Leipzig 2011. Merkur GmbH, Leipzig, 2011; S. 61–62.