

Probleme und Möglichkeiten der Nutzung einzeltierbezogener Daten in Milchviehbetrieben

Christian Ammon, Joachim Spilke

Landwirtschaftliche Fakultät, AG Biometrie und Agrarinformatik
Martin-Luther-Universität Halle-Wittenberg
Ludwig-Wucherer-Str. 82-85
06108 Halle
christian.ammon@landw.uni-halle.de
joachim.spilke@landw.uni-halle.de

Abstract: Decision Support Systems, here in dairy cows, ask for certain database requirements. This article shows some of the problems that occurred while working with data from four different dairy farms as well as shows methods and approaches used by the authors for further processing of that data.

1 Einleitung

In Milchviehbetrieben stehen durch immer weiterentwickelte (Melk-)Technik zunehmend einzeltierbezogene Daten zur Verfügung. In Herdenmanagementprogrammen sind diese Daten zu Informationen über das Einzeltier, die Tiergruppe oder die Herde aufzuwerten. Dazu sind Entscheidungsunterstützungssysteme erforderlich, die die Aufmerksamkeit des Personals auf bestimmte Tiere lenken und Basis für Managemententscheidungen sind. Um die Funktion der Entscheidungsunterstützung erfüllen zu können, müssen allerdings einige Anforderungen an die Datenbasis gestellt werden.

Im vorliegenden Beitrag werden daher zunächst die Ansprüche an die Datenbereitstellung untersucht. Anhand von Beispielen wird gezeigt, dass diese nicht immer gegeben sind. Anhand von Analyseergebnissen aus vier Betrieben mit unterschiedlichen Melksystemen werden die beschriebenen Probleme dargestellt. Der zweite Teil umfasst die Anforderungen, die für die Datennutzung erfüllt sein müssen. Die Kompatibilität von Daten aus verschiedenen Quellen muss gewährleistet sein, wenn unterschiedliche Merkmale in einem Entscheidungsunterstützungssystem miteinander kombiniert werden. Dies gilt nicht nur für die hier nicht weiter betrachteten Datenformate, sondern auch für die Häufigkeit von Messungen, Datenermittlungen und -übertragungen. Die derzeit von den Autoren verwendeten Methoden und Ansätze zur Datennutzung umfassen gemischte lineare Modelle, Fuzzy Logic und Exponentially Weighted Moving Average (EWMA) Control-Charts. Deren Einsatzmöglichkeiten und Einordnung sowie die notwendigen Anpassungen zur Verwendung im Informationssystem eines Milchviehbetriebs werden näher erläutert.

2 Datenbereitstellung

Erforderlich sind eindeutig einem Tier zuzuordnende konsistente und kontinuierlich vorhandene Daten. Bei den vier betrachteten Betrieben handelt es sich um zwei Produktionsbetriebe und zwei Versuchsbetriebe mit unterschiedlichen Herdengrößen und Melkrythmen. Die Daten der beiden Produktionsbetriebe und eines der Versuchsbetriebe stammen aus der normalen Produktion, während im vierten Betrieb zwei zeitlich ineinander übergehende Fütterungsversuche die Datenbasis liefern. Tab. 1 gibt einen Überblick über die vorliegenden Daten.

	Betrieb 1	Betrieb 2	Betrieb 3	Betrieb 4
Zeitraum (Tage)	396	806	1475	914
Gesamtanzahl Tiere	532	394	77	661
Durchschnittliche Anzahl Tiere/Tag	238 (Aufstockung von 120 auf 340 während des Zeitraums)	130 (Aufstockung von 120 auf 180 während des Zeitraums)	23 (schwankende Anzahl Versuchstiere)	322
Datensätze	82881	236706	34586	266696
Durchschnittliche Laktation	2.17	1.92	2.25	2.59
Durchschnittliche tägliche Milchleistung (kg)	30	32	25	34
Melkungen/Tag	3 (58%) bzw. 2 (42%)	2.94 (AMS)	2	2
Ausfalltage	48	199	0	85
Datenabdeckung des Zeitraums	87.9%	75.3%	100%	90.7%
Fehlzuordnungen	28988 (35%)	149 (0%)	0 (0%)	~2100 (0.8%)

Tabelle 1: Übersicht der Datenbasis der vier Betriebe

Der erste Betrieb ist ein Produktionsbetrieb mit verschiedenen Leistungsgruppen. Vier Gruppen werden dreimal, fünf Gruppen zweimal täglich gemolken. Hier beginnt ein „Verständigungsproblem“ zwischen Melksystem und Herdenmanagementsystem: während vom Melksystem beispielsweise nur nach Melkungen geordnet wird, sortiert das Herdenmanagementsystem nach Melkzeiten, so dass z.B. Gemelk 2 im Prozessrechner nicht Gemelk 2 im Managementprogramm entspricht. Zusätzlich treten bisweilen Diskrepanzen zwischen den Milchmengen einerseits je Tier und Tag und andererseits auch je Einzelgemelk auf. Offensichtlich kommt es auch zu Fehlzuordnungen von Gemelken durch das Melkpersonal, so dass Milchmengen von mehreren Kühen zu einem Gemelk zusammengefasst werden. Diese Fehlzuordnungen lassen sich identifizieren, indem man die tatsächliche Anzahl der Gemelke mit Milchmengen größer Null mit der durch die Leistungsgruppe vorgegebenen Anzahl Melkungen je Tag abgleicht. So finden sich Datensätze, bei denen z.B. ein Gemelk mit sehr hoher Milchmenge (30 und mehr kg) und zwei Gemelke mit Milchmenge 0 vorliegen. Es ist aber nicht ersichtlich, ob der Wert des ersten Gemelks nur von diesem Tier stammt und die Tagesleistung dieses Tieres zusammenfasst, oder ob beim ersten Melken Gemelke von drei (oder mehr) Tieren

erfasst wurden, während der registrierte Transponder noch der eines anderen Tieres war und vom Personal nicht manuell weitergeschaltet wurde. Da dies recht häufig geschieht (vgl. Fehlzuordnungen in Tab. 1) und auch nachträglich keine korrekte Zuordnung mehr erfolgen kann, sind diese Daten für einzeltierbezogene Zwecke nicht verwendbar.

Vom zweiten Betrieb, ebenfalls einem Produktionsbetrieb, jedoch mit automatischem Melksystem (AMS), ist durchschnittlich etwa jeder vierte Tag ein „Datenausfalltag“. Neben der Milchleistung werden hier noch weitere, viertelbezogene Merkmale erfasst, wobei gelegentlich Datenübertragungsfehler auftreten: Es fehlen bei etwa 16% der Datensätze die Messwerte für Leitfähigkeit der Milch, bei 3.4% der Datensätze fehlt die Melkdauer, während 0.5% der Datensätze trotz vorhandener Milchmenge eine Melkdauer von 0s aufweisen. Dies kann auch ein Hinweis auf Fehlbuchungen sein, die durch Erkennung von Transpondern von Kühen außerhalb der Melkbox entstehen. Ersichtlich wird dies durch sehr kurze festgehaltene Melkzeiten bei sehr hohen Milchflusswerten. Für weitere Anwendungen war es daher erforderlich, den Zeitraum der verwendeten Daten so zu wählen, dass möglichst wenige Fehltag auftraten. Der Datensatz des dritten Betriebs umfasst nur die Messwerte der in zwei Fütterungsversuchen beteiligten Tiere. Die Werte werden erfasst, sobald eine Kuh abgekalbt hat und die neue Laktation beginnt; nachdem die Laktation vorüber ist verlassen sie meistens den Versuch. Von nur 23 aller 77 Tiere sind Messwerte von zwei Laktationen vorhanden, bei nur einem sind es drei Laktationen. Dadurch ergibt sich zu Beginn und gegen Ende der Versuchsdauer eine nur geringe Anzahl Versuchstiere, so dass bei Schätzung der Testtageeffekte größere Schwankungen auftreten. Tier- und Testtageeffekt kann in der statistischen Auswertung schwer getrennt werden. Daher wurde für weitere Berechnungen der Zeitraum der verwendeten Daten eingeschränkt, um eine geeignete Mindestanzahl von Tieren zur Verfügung zu haben. Beim vierten Betrieb handelt es sich um einen Versuchsbetrieb. Auch hier tritt das Problem von Fehlzuordnungen von Gemelken zu einzelnen Tieren auf, allerdings nur bei etwa 0.8% der Datensätze. Mangels weiterer Daten aus dem Herdenmanagementsystem ist die Anzahl der Fehlzuordnungen nicht eindeutig festzulegen; die in Tab. 1 befindlichen Angaben wurden durch visuelle Kontrolle der einzeltierspezifischen Milchleistungskurven auf Ausreißerwerte erlangt. Da nur Milchleistungsdaten zur Verfügung standen, wurde dieser Datensatz für weitere Bearbeitung zunächst zurückgestellt.

Es ist festzuhalten, dass konsistente Daten im Sinne einer eindeutigen Zuordenbarkeit für die Verwendung von einzeltierspezifischen Daten unumgänglich sind. Lückenhafte Daten sind zu vermeiden, da sie die Verwendung von Algorithmen, die sowohl auf aktuellen als auch auf zurückliegenden Daten basieren, entweder erschweren oder unmöglich machen.

3 Konsequenzen für die Datennutzung in Milchviehbetrieben

Zunächst wurden die Daten aus den Betrieben 2 und 3 verwendet, um ein gemischtes lineares Modell für die Tagesmilchleistungen anzupassen [AS04]. Mit diesem Modell lassen sich einzeltier- und herdenbezogen laktationsspezifische Laktationskurven schätzen, es können Testtageeffekte geschätzt und Milchleistungen kurzfristig vorhergesagt

werden. Diese Berechnungen gehen in der Regel über die Möglichkeiten eines Milchviehbetriebs hinaus und müssen ausgelagert werden; daraus wird der Bedarf eines Standardaustauschformats ersichtlich. Die Vorhersagewerte bilden auch die Grundlage für ein Fuzzy Logic-Modell zur Erkennung von Leistungsabweichungen [AS05]. Dabei wird ebenfalls die Notwendigkeit von kontinuierlich vorhandenen Daten je Tier offensichtlich. Da sich die Betrachtung absoluter Differenzen zwischen tatsächlicher und vorhergesagter Milchleistung wegen möglicher Über- und Unterschätzung der tatsächlichen Milchleistung verbietet, müssen relative, den Trend über mehrere Tage hinweg wiedergebende Abweichungen verwendet werden. Fehlende Daten führen entweder zu ungenaueren Ergebnissen oder dazu, dass eine Inputvariable des Modells fehlt und somit kein Outputwert für diesen Tag berechnet werden kann. Ein weiteres Hindernis bei der Verwendung eines Fuzzy Logic-Modells kann die unterschiedliche Verfügbarkeit von Inputvariablen sein. Bei einem Modell zur Erkennung von Euterkrankheiten sind z.B. Leitfähigkeit der Milch (je Melkung verfügbar) und Zellzahl der Milch (wöchentlich verfügbar) als Inputvariablen vorstellbar. Da sich die Zellzahl jedoch nur einmal wöchentlich als Input ändert, ist der Wert für eine aktuelle Entscheidungshilfe fraglich.

EWMA Control-Charts (vgl. z.B. [En05]) erwiesen sich als nicht besonders geeignet zur Erkennung von Leistungsabweichungen. Bei starker Gewichtung der aktuellen Daten erhält man ein sehr nahe an den tatsächlichen Leistungen befindliches Chart, so dass nur sehr extreme Abweichungen feststellbar sind, während stärkere Gewichtung von weiter zurückliegenden Daten zu einer schlechten Anpassung an die Laktationskurve führt. Die Verwendung für Merkmale mit weniger stark schwankenden Messwerten während einer Laktation, wie z.B. Leitfähigkeit der Milch, ist jedoch denkbar.

Als Konsequenz für die Datennutzung in Milchviehbetrieben ergibt sich der Bedarf für sichere, prozessnahe Datenerfassung, kompatible Datenaustauschformate vom Prozessrechner über Herdenmanagementprogramm bis zur Datenbank eines Dienstleisters und Wahl der geeigneten Verarbeitungsmethode in Abhängigkeit vom betrachteten Merkmal.

Literaturverzeichnis

- [AS04] Ammon, C.; Spilke, J.: Vergleich von Fixed- und Random-Regression Modellen bei verschiedenen Funktionsansätzen für Laktationskurven zur Vorhersage von Milchleistungen. In: Integration und Datensicherheit - Anforderungen, Konflikte und Perspektive: Referate der 25. GIL Jahrestagung, 8.-10. September 2004 in Bonn, Germany. Bonn: Ges. für Informatik, 2004; S. 149-152.
- [AS05] Ammon, C.; Spilke, J.: Comparison of fixed- and random-regression models using different functional approaches of lactation curves for milk yield forecasts. In (Boaventura Cunha, J., Morais, R., Hrsg.): Proceedings EFITA/WCCA Joint Conference Vila Real 2005, Univ. de Trás-os-Montes e Alto Douro Vila Real, 2005; S. 630-635.
- [En05] Engler, J. et al.: Control charts applied to individual sow farm analysis. In (Cox, S., Hrsg.): Precision Livestock Farming '05, Proceedings ECPLF Uppsala 2005, Wageningen Academic Publishers, 2005; S. 319-325.