ISSN: 1563-2504
ISBN: 978-92-95014-22-0

THE GLOBAL STANDARD
FOR LIVESTOCK DATA

Network. Guidelines. Certification.

ICAR Technical Series no. 26

SHARE. CONNECT. TRANSFORM. SHARING SOLUTIONS ON DIGITAL TRANSFORMATION, ANIMAL WELFARE AND ENVIRONMENTAL SUSTAINABILITY TO SUPPORT GLOBAL FOOD SECURITY

Proceedings of the 45th ICAR Annual Conference
held in Montréal, Quebec, Canada
30 May - 3 June 2022

Editors: A.-M. Christen, D. Lefebvre, F. Miglior, R. Cantin, M. Burke, and C. Mosconi

The designations employed and the presentation of material in this information product do not imply the expression of any opinion whatsoever on the part of ICAR concerning the legal or development status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. The mention of specific companies or products of manufacturers, whether or not these have been patented, does not imply that these have been endorsed or recommended by ICAR in preference to others of a similar nature that are not mentioned. The views expressed in this information product are those of the author(s) and do not necessarily reflect the views or policies of ICAR.

ICAR encourages the use, reproduction and dissemination of material in this information product. Except where otherwise indicated, material may be copied, downloaded and printed for private study, research and teaching purposes, or for use in non-commercial products or services, provided that appropriate acknowledgement of ICAR as the source and copyright holder is given and that ICAR's endorsement of users' views, products or services is not implied in any way.

All the manuscripts are published under the responsibility of each Author. ICAR is in no way responsible for the opinion expressed by each Author
.

Title of the Series: ICAR Technical Series

Title of the Volume: Share. Connect. Transform.
Sharing solutions on digital transformation, animal welfare and environmental sustainability to support global food security

Editors: A.-M. Christen, D. Lefebvre, F. Miglior, R. Cantin,
M. Burke, and C. Mosconi

Publisher: ICAR, Arthur van Schendelstraat 650, 3511 MJ Utrecht,
The Netherlands

ISSN: 1563-2504
ISBN: 978-92-95014-22-0

Responsible for the ICAR Technical Series: Cesare Mosconi

ICAR and the local organisers wish to thank the following companies and organisations for sponsoring the 2022 ICAR/Interbull Conference

Gold sponsors

BENTLEY INSTRUMENTS

Caisley INTERNATIONAL GMBH

DAIRY FARMERS OF CANADA
LES PRODUCTEURS LAITIERS DU CANADA

FOSS

PARTENARIAT CANADIEN pour l'AGRICULTURE
Canada
Québec

MERCK Animal Health
Allflex Livestock Intelligence™

Mtech Digital Solutions

uniform agri

Silver and Bronze sponsors

Silver Sponsors

DATAMARS Livestock

LELY

SEMEX

Bronze Sponsors

DNA DIAGNOSTIC

illumina

AgSource
vas

KPI DIGITAL

Exhibitors at the Conference

Bentley Instruments

Caisley

Dairy Farmers of Canada

Datamars Livestock

Eurofins

European Milk Recording

FOSS

IDEXX

Medria Solution / GenoCells / Eliance / Sayca

MSD Animal Intelligence

Mtech Digital Solutions

Safeguard Biosystems

Uniform-Agri

THE GLOBAL STANDARD
FOR LIVESTOCK DATA

Network. Guidelines. Certification.

ICAR Technical Series no. 26

SHARE. CONNECT. TRANSFORM.

SHARING SOLUTIONS ON DIGITAL TRANSFORMATION, ANIMAL WELFARE AND ENVIRONMENTAL SUSTAINABILITY TO SUPPORT GLOBAL FOOD SECURITY

Proceedings of the 45th ICAR Annual Conference
held in Montréal, Quebec, Canada
30 May - 3 June 2022

Editors: A.-M. Christen, D. Lefebvre,
F. Miglior, R. Cantin,
M. Burke, and C. Mosconi

November 2022

Introduction

Bonjour Madame, Bonjour Monsieur,

Lactanet was proud to host the 45th annual ICAR conference in conjunction with the Interbull meeting. The uncertainty around the worldwide management of COVID-19 regulations was challenging for the organizing committee, but three years after our last in-person event, we were delighted that 363 participants registered for the conference; 297 attended the meeting in person, while 66 participated virtually from home. Our guests, coming from 42 countries, had the opportunity to immerse themselves in Montréal, this vibrant, multicultural metropolis, the business hub of the French-speaking province of Québec, and also discover the quaint setting of «Vieux-Montréal» and its charming little streets.

Share. Connect. Transform.

The topics, the relevance and the quality of the presentations demonstrated that we, as animal improvement experts and scientists, are taking action and looking for innovative and practical solutions to help the livestock sector face the upcoming environmental challenges. In total, 82 abstracts were submitted and distributed amongst eight themes. The «Advanced Analytics for Adding Value to Livestock Data» session was particularly popular, highlighting that artificial intelligence and data analytics are now integral to the dairy science and animal agriculture toolbox.

Our planet is facing alarming climate change. The livestock sector has been called upon to make their contribution to limiting global warming and will need to adapt to the changing climate. International collaboration is vital for sharing and implementing global actions that will bring about the necessary changes to meet these challenges. Besides exploring solutions through research, the livestock sector urgently needs tools and information on how to select the next generation of animals to be more resilient and respond better to these challenges. The presentations on feed efficiency, data analytics and the use of sensors and advanced automated systems in our routine farm practices, showed us that we can develop solutions to help farm managers monitor and predict changes better, allowing them to optimize their operations. Beyond the primary objective of feeding people, animal agriculture now concerns itself with protecting our resources, using the right inputs, and connecting with consumers to meet their demands and expectations for nutritious, safe, affordable and sustainable food. Integrating multidisciplinary sciences is our challenge, along with being open to new opportunities that could transform our sector into the sustainable, efficient, and resilient one that will feed the world.

The success of such a conference comes with the support of its valued partners

We would like to extend our heartfelt thanks and appreciation to the sponsors and exhibitors who supported the organization of the conference. Their generous support allowed the ICAR and Interbull communities to meet and share their knowledge and expertise to advance the livestock sector. We hope that the Montréal conference has brought about collaboration and business opportunities. Acknowledgement also goes out to the ICAR Secretariat team for their support and kind help with the general organization, and to the AgoraOpus3 team, who supported us from the beginning of this great adventure to the final day.

Finally, we would like to thank the Quebec Ministry of Agriculture, Fisheries and Food (MAPAQ) and the Canadian Agricultural Partnership funding program for their support.

Daniel Lefebvre
Chief Operations Officer for Lactanet and President of ICAR

PARTENARIAT
CANADIEN pour
l'AGRICULTURE

Canada Québec

« Ce projet est financé en vertu de l'Accord Canada-Québec de mise en œuvre du Partenariat canadien pour l'agriculture. Ensemble, le gouvernement fédéral et le gouvernement du Québec ont investi 293 millions de dollars répartis sur une période de 5 ans, soit de 2018 à 2023. Cet accord appuie des initiatives stratégiques qui aideront les secteurs à croître, à innover et à prospérer».

Table of content

Update ICAR Feed and Gas Working Group: Guidelines and international collaboration for genetic evaluation

B. Gredler-Grandl[1], C. Baes[7], Y. de Haas[1], R. Finocchiaro[2], P. Garnsworthy[3], N. Krattenmacher[4], J. Lassen[5], J. Pryce[6] and R. Veerkamp[1]

[1]Wageningen University and Research, Animal Breeding and Genomics Centre, P.O. Box 338, 6700 AH Wageningen, The Netherlands
Corresponding Author: birgit.gredler-grandl@wur.nl
[2]Associazione Nazionale Allevatori della Razza Frisone, Brune e Jersey Italiana),
Via Bergamo 292, 26100 Cremona, Italy
[3]School of Biosciences, University of Nottingham, Sutton Bonington Campus,
Loughborough LE12 5RD, United Kingdom
[4]Institute of Animal Breeding and Husbandry, Christian-Albrechts-University,
Hermann-Rodewald-Str. 6, 24098 Kiel, Germany
[5]VikingGenetics, Ebeltoftvej 16, 8960 Randers, Denmark
[6]School of Applied Systems Biology, La Trobe University,
Bundoora, Victoria 3083, Australia
[7]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences,
University of Guelph, Guelph, ON N1G 2W1, Canada

Abstract

Feed efficiency and methane emissions are important trait groups with regard to environmental impact, sustainability and efficiency of dairy cattle production. The ICAR Feed and Gas working group was established as a permanent ICAR working group in 2015 to support ICAR's key role of providing guidelines on animal recording to support farmers in daily management decisions and to deliver validated information for genetic evaluations. The working group developed its first guidelines in 2020. Current priorities of the working group are to update and extend these guidelines and facilitate and evaluate possibilities for international genetic evaluation.

Keywords: Feed intake, methane emission, guidelines,, international collaboration, genetic evaluation.

Introduction

The ICAR Feed and Gas working group was established as a permanent ICAR working group in 2015 to support ICAR's key role of providing guidelines on animal recording to support farmers in daily management decisions and to deliver validated information for genetic evaluation. The main objectives of the group are:

- To update and extend guidelines for recording dry matter intake and methane emissions in cattle, sheep and goats worldwide.
- To provide a forum to ensure international exchange of knowledge and project results for feed intake and methane emission and coordinate international collaboration in research and development.
- To conduct and report periodic international surveys

- To elaborate challenges and possibilities for international genetic evaluation for feed efficiency and methane emission.

Update guidelines

The ICAR Feed and Gas Working Group developed guidelines for recording and genetic evaluation of individual feed intake and methane emissions in 2020. There is a need to clearly define and standardize recordings for feed intake and methane emissions. The group works on updates of the guidelines including results of new developments in recording techniques, trait definitions and trait evaluation for feed intake and methane emissions. For instance, precision livestock farming provides new solutions to record individual feed intake by using 3D cameras to get reliable estimates of individual daily feed intake (e.g. Lassen et al., 2018, Bezen et. al. 2020). This technique is applied by research and industry organisations and first results and experiences are available. In particular, for individual methane emissions data new research results became available in the last two years. Guidelines will be reviewed to account for evolving science in the field, and practical tips on how to edit raw data to produce a reliable phenotype record for management purposes or further genetic analyses.

Several countries, including Australia, United States of America, Denmark, Norway, Finland, New Zealand, The Netherlands, United Kingdom, and Canada have implemented a genetic evaluation of feed efficiency in the recent years (summarized by e.g. Brito et al. 2020 and Stephansen et al., 2021). Those countries are invited to provide descriptions of national genetic evaluations for feed efficiency. This information is provided in the genetic evaluation form of Interbull (GE Form). Here, practical information about data editing steps and models used for genetic evaluations will be provided. In future, the GE forms will be incorporated in the newly developed PREP database by the Interbull Centre. The PREP database is a platform developed by the Interbull Centre to enable users to upload descriptive information regarding performance recording, national genetic evaluation systems, and publication policies. The Feed & Gas working group and the Interbull Centre will work together closely to set up the PREP database for feed efficiency traits.

International collaboration for genetic evaluation

The widespread use of genomic information has enabled selection for scarcely recorded traits such as individual feed intake and methane emissions. However, to achieve desirable reliabilities of genomic breeding values for feed efficiency traits that are developed from feed intake data, a large reference population of 30,000 animals is suggested (Brito et. al. 2021). de Haas et. al. (2021) estimated that recording individual methane emissions on an average of 150 cows across 100 farms for a minimum period of 2 years is needed to achieve a desirable reliability of genomic breeding values for methane. Phenotyping several thousands of individuals for difficult to measure traits is still challenging for single countries. International collaborations to share and exchange phenotypes are of extreme importance in this regard. Here, we mention three initiatives to share data across countries.

The **Global Dry Matter Initiative** (gDMI, de Haas et. al. 2014) was the first research initiative to collate feed intake records across countries in an international reference population to estimate genomic breeding values for dry matter intake (DMI). Fifteen partners (both science and industry) from nine countries worldwide collaborated to form a reference population of around 10,000 phenotyped and 6,000 genotyped animals. The outcome of the collaboration showed clear benefits by improving reliabilities of genomic breeding values for DMI for most countries (de Haas et. al. 2015).

The **Resilient Dairy genome Project** (RDGP; http://www.resilientdairy.ca) is a large-scale applied research project funded by Genome Canada with the overall aim to provide genomic tools for selection of more resilient dairy cattle. The focus of RDGP is to build a large international reference population for feed efficiency and methane emission. The RDGP database currently holds 12,687 feed intake and 3,093 methane emission records of individual animals from seven partners. Within the RDGP all data are accessible by all partners to be used in their own national genetic evaluations.

Re-Livestock – Facilitating Innovations for Resilient Livestock Farming Systems is a large-scale international project to be started in September 2022 addressing challenges of the work programme HORIZON-CL6-2021-CLIMATE-01-06. The overall objective of Re-Livestock is to adopt and apply innovative practices across-scale (animal, herd/farm, and sector) to reduce greenhouse gas emissions of livestock and increase resilience of the livestock sector. A work package is dedicated to investigating the contribution of breeding to a reduced impact of livestock on climate change mitigation and to the adaptation of livestock to climate change. Phenotypes and genotypes of around 13,000 cows (dairy and beef) will be merged across countries to perform multi-trait genomic predictions.

Sharing and exchange of scarce and expensive-to-record phenotypes for genetic evaluation has been very successful within research projects (see initiatives mentioned above) for feed efficiency (and methane emissions). In commercial applications, sharing data is still sensitive, and often data ownership issues and commercial interests hinder collaboration across countries. However, possibilities for the future are to apply methods, such as meta-analysis of national genomic evaluation results or the Interbull GMACE or SNP MACE methods (Jighli et. al., 2019) to increase the reliability of genomic predictions for feed efficiency traits and methane emissions.

References

Bezen, R., Y Edan, I. Halachmi 2020 Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms Computers and Electronics in Agriculture 172:105345

Brito, L. F., H. R. Oliveira, K. Houlahan, P. A. S. Fonseca, S. Lam, A. M. Butty, D. J. Seymour, G. Vargas, T. C. S. Chud, F. F. Silva, C. F. Baes, A. Cánovas, F. Miglior, and F. S. Schenkel. 2020. Genetic mechanisms underlying feed utilization and implementation of genomic selection for improved feed efficiency in dairy cattle. Can. J. Anim. Sci. 100:587–604. doi:10.1139/cjas-2019-0193.

Brito, L. F., N. Bedere, F. Douhard, H. R. Oliveira, M. Arnal, F. Peñagaricano, A. P. Schinckel, C. F. Baes, and F. Miglior. 2021. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. Animal. 100292. doi:10.1016/j.animal.2021.100292.

de Haas, Y., J.E. Pryce, M.P.L. Calus, I. Hulsegge, D.M. Spurlock, D. Berry, E. Wall, P. Løvendahl, K.A. Weigel, K. Macdonald, F. Miglior, N. Krattenmacher, R.F. Veerkamp. 2014. Genomic Predictions for Dry Matter Intake Using the International Reference Population of gDMI. Interbull Bulletin 48, 94-99.

de Haas, Y., J. E. Pryce, M. P. L. Calus, E. Wall, D. P. Berry, P. Løvendahl, N. Krattenmacher, F. Miglior, K. Weigel, D. Spurlock, K. A. Macdonald, B. Hulsegge, and R. F. Veerkamp. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North America, and Australasia. J. Dairy Sci. 98:6522–6534. doi:10.3168/jds.2014-9257.

de Haas, Y., R. F. Veerkamp, G. de Jong, and M. N. Aldridge. 2021. Selective breeding as a mitigation tool for methane emissions from dairy cattle. Animal. 100294. doi:10.1016/j.animal.2021.100294.

Lassen, J., J. R. Thomasen, R. H. Hansen, G. G. B. Nielsen, E. Olsen, P. R. B. Stentebjerg, N. W. Hansen, and S. Borchersen. 2018. Individual measure of feed intake on inhouse commercial dairy cattle using 3D camera technology, Proceedings of the World Congress on Genetics Applied to Livestock Production, vol. Technologies   Novel Phenotypes, p. 635, 2018.

Jighly, A., H. Benhajali, Z. Liu and M.E. Goddard. 2019. SNPMace -A meta-analysis to estimate SNP effects across countries. Interbull Bulletin 55, 107-115.

Stephansen, R. B., J. Lassen, J. F. Ettema, L. P. Sørensen, and M. Kargo. 2021. Economic value of residual feed intake in dairy cattle breeding goals. Livest. Sci. 253:104696. doi:10.1016/j.livsci.2021.104696.

Measuring individual carbon dioxide emissions as a proxy for feed efficiency on dairy farms – preliminary results

I. Fodor[1], N. Ogink[2], F. de Jong[3] and Y. de Haas[1,4]

[1]Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338,
6700 AH Wageningen, The Netherlands
Corresponding Author: istvan.fodor@wur.nl
[2]Livestock and Environment, Wageningen University and Research,
P.O. Box 338, 6700 AH Wageningen, The Netherlands
[3]Dairy Campus, Wageningen University and Research, Boksumerdyk 11,
8912 CA Leeuwarden, The Netherlands
[4]Current: Animal Health and Welfare, Wageningen University and Research,
P.O. Box 338, 6700 AH Wageningen, The Netherlands

Abstract

The costs and practical challenges of measuring individual feed efficiency on dairy farms have hampered its use both in farm management and genetic selection. The concept of residual carbon dioxide (RCO_2) is similar to the index of residual feed intake (RFI), and used as a proxy, it could potentially enable ranking cows based on feed efficiency. In this study, we aimed to assess the relationship between RCO_2 and RFI in mid-lactation (115 to 175 days in milk), using data of 313 cow-per-treatment observations from five experiments with individual feed intake records. Carbon dioxide production was measured by GreenFeed units. Residual CO_2 (kg/day) and RFI (kg of dry matter/day) were estimated using a mixed model approach. Three cow groups (high/mid/low) of equal sizes were created both for RCO_2 and RFI. Cows in the high-RCO_2 group produced 1.92 kg/day (95% CI: 1.78 – 2.06, $P < 0.0001$) more carbon dioxide than their low-RCO_2 herd mates. At the same time, high-RCO_2 cows had 1.31 kg/day (95% CI: 0.95 – 1.67) higher RFI compared to their low-RCO_2 counterparts ($P<0.0001$). The overall classification accuracy based on RCO_2, using RFI as a reference, was 48.2%, however, inefficient (i.e. high RFI) cows could be identified with 59.0% accuracy. In conclusion, relevant differences were found between RCO_2 groups in terms of feed efficiency. Therefore, routinely collecting individual carbon dioxide emissions seems to be a promising tool to record feed efficiency on a large scale.

Keywords: dairy cattle, feed efficiency, residual feed intake, carbon dioxide, GreenFeed.

Introduction

Dairy production has to meet the increasing demand for milk and dairy products of the growing world population. Improved feed efficiency of dairy cows has large potential to reduce the environmental impact of this growth in multiple ways. First of all, land requirements of feed production can be decreased by 6.7% per cow via reducing feed intake by one standard deviation, at the same level of energy-corrected milk yield (Bell *et al.*, 2012). Moreover, traditional selection for improved feed efficiency can decrease methane emissions per kg of fat- and protein corrected milk by 26% over a selection period of 10 years (de Haas *et al.*, 2011). More feed efficient cows also produce less manure, thereby reducing the amount of methane and nitrous oxide released into the atmosphere (Connor, 2015). Farmers also benefit from improved feed efficiency through reduced feed costs, which represent more than 50% of the total cost of milk production (European Commission, 2020). As a moderately heritable trait, feed efficiency could be improved by selective breeding, although, the costs and practical limitations of recording dry matter intake (DMI) limit selection (Seymour *et al.*, 2019).

The concept of residual carbon dioxide (RCO_2) is similar to the index of residual feed intake (RFI), and used as a proxy alone or in combination with other easily available parameters, it could potentially enable ranking cows based on feed efficiency (Huhtanen *et al.*, 2021). The concept of RCO_2 was originally developed using data from respiration chambers, but using sensors with highly repeatable carbon dioxide measurements (e.g. GreenFeed), the approach could be used on the farms, as well. In this study, we aimed to assess the relationship of RCO_2 and RFI in mid-lactation on a dairy farm.

Materials and methods

Data

The data of 313 cow-per-treatment observations from five experiments, carried out at Dairy Campus of Wageningen University and Research (Leeuwarden, The Netherlands), were used in the analyses. Dry matter intake (kg/day), energy-corrected milk (kg/day), daily average body weight (kg), and carbon dioxide production (kg/day) were collected. Carbon dioxide production was measured by GreenFeed (C-Lock, Inc., Rapid City, SD) units. Dry matter intake from GreenFeed bait was not taken into account. Records between 115 and 175 days in milk were retained for the analyses, to minimize the possible effects of changes in energy balance, and because RFI measured in this period is highly correlated with the average RFI over the whole lactation (Martin *et al.*, 2021).

Statistical analysis

Residual CO_2 and RFI were obtained from mixed-effects models, accounting for energy-corrected milk and metabolic body weight (MBW, i.e. body weight$^{0.75}$) as fixed effects, and treatment and experiment as random effects. Cows were classified into three equal-sized groups (high/mid/low, n = 104 – 105 each) based on RCO_2 and RFI. The resulting groups were compared using linear models, followed by Tukey's post hoc tests for pairwise comparisons.

Results and discussion

The descriptive statistics of the parameters are shown in Table 1. The level of dry matter intake, energy-corrected milk production, and the body weight of cows were comparable to those of Huhtanen *et al.* (2021).

Table 1. Descriptive statistics of the analysed parameters.

Parameter	Mean	SD
Dry matter intake (kg/day)	20.5	3.5
Carbon dioxide production (kg/day)	13.7	1.3
Energy-corrected milk (kg/day)	31.7	5.3
Metabolic body weight (kg)	130.8	10.0

Significant differences in RCO_2 were observed between RCO_2 groups (Table 2). The low-RCO_2 group produced on average 1.92 kg/day less carbon dioxide than the high-RCO_2 group, at the same level of energy-corrected milk yield and metabolic body weight ($P < 0.0001$). Feed efficiency, expressed in RFI, differed significantly between RCO_2 groups. The low-RCO_2 group consumed 1.31 kg less feed per day than the high-RCO_2 group, at the same energy-corrected milk production and metabolic body weight ($P < 0.0001$). Our results support the findings of Huhtanen *et al.* (2021), who found similar differences between low- and high-RCO_2 groups both in terms of RCO_2 and RFI.

Residual CO_2 and RFI groups are cross-tabulated in Table 3. Overall, 48.2% of the cows were correctly classified based on RCO_2, using RFI as reference, however, inefficient (i.e. high-RFI) cows could be identified with 59.0% accuracy. Inefficient cows were rarely misclassified as being efficient (16.2%), and vice versa (14.4%). This supports the potential of residual carbon dioxide to be used as a proxy for feed efficiency on dairy farms.

On the other hand, care should be taken when evaluating RCO_2, because energy balance influences RCO_2. For example, cows in negative energy balance mobilize body fat reserves to produce milk fat, which process does not generate carbon dioxide, consequently, these cows can be erroneously considered efficient. Therefore, carbon dioxide measurements should be performed in mid-lactation, when the probability of such misclassification is low (Huhtanen *et al.*, 2021). Alternatively, changes in body condition or milk composition can be followed to obtain information on energy balance (Friggens *et al.*, 2007; Thorup *et al.*, 2018).

Table 2. Differences in residual carbon dioxide production (RCO_2) and residual feed intake (RFI) between RCO_2 groups.

	RCO_2			RFI		
RCO_2 group	**Difference (kg/day)**	**95% CI**	**P-value**	**Difference (kg/day)**	**95% CI**	**P-value**
Low vs. high	-1.92	-2.06 – -1.78	<0.0001	-1.31	-1.67 – -0.95	<0.0001
Mid vs. high	-1.02	-1.16 – -0.88	<0.0001	-0.89	-1.25 – -0.53	<0.0001
Low vs. mid	-0.90	-1.04 – -0.76	<0.0001	-0.42	-0.78 – -0.06	0.0168

Table 3. Number (% in parentheses) of cows by residual carbon dioxide (RCO_2) group, using residual feed intake (RFI) groups as reference.

	RCO_2 group		
RFI group	**High (n = 105)**	**Mid (n = 104)**	**Low (n = 104)**
High (n = 105)	62 (59.0)	26 (24.8)	17 (16.2)
Mid (n = 104)	28 (26.9)	39 (37.5)	37 (35.6)
Low (n = 104)	15 (14.4)	39 (37.5)	50 (48.1)

Conclusion

We confirmed the utility of residual carbon dioxide as a proxy for feed efficiency on dairy farms. For future use in practice, carbon dioxide measurements should either be performed in mid-lactation, or preferably, energy balance of the animals should be estimated in parallel. Individual carbon dioxide measurements offer potential to distinguish efficient from inefficient cows on dairy farms, without the need to measure individual dry matter intake.

Acknowledgement

This study was supported by the TKI Agri and Food project LWV19143 and the partners Melkveefonds and Connecterra.

References

Bell, M.J., E. Wall, G. Russell, G. Simm and A.W. Stott, 2012. The effect of improving cow productivity, fertility, and longevity on the global warming potential of dairy systems. J. Dairy Sci. 94: 3662–3678.

Connor, E.E., 2015. Invited review: Improving feed efficiency in dairy production: challenges and possibilities. Animal 9(3): 395–408.

de Haas, Y., J.J. Windig, M.P.L. Calus, J. Dijkstra, M. de Haan, A. Bannink and R.F. Veerkamp, 2011. Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. J. Dairy Sci. 94: 6122–6134.

European Commission, 2020. EU agricultural and farm economics briefs. No. 18. EU milk margin index estimate up to 2020. https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/farming/documents/agri-farm-economics-milk-margin-estimate-2020_en.pdf

Friggens, N.C., C. Ridder and P. Løvendahl, 2007. On the use of milk composition measures to predict the energy balance of dairy cows. J. Dairy Sci. 90:5453–5467.

Thorup, V.M., M.G.G. Chagunda, A. Fischer, M.R. Weisbjerg and N.C. Friggens, 2018. Robustness and sensitivity of a blueprint for on-farm estimation of dairy cow energy balance. J. Dairy Sci. 101:6002–6018.

Huhtanen, P., A. Bayat, P. Lund and A. Guinguina, 2021. Residual carbon dioxide as an index of feed efficiency in lactating dairy cows. J. Dairy Sci. 104: 5332–5344.

Seymour, D.J., A. Cánovas, C.F. Baes, T.C.S. Chud, V.R. Osborne, J.P. Cant, L.F. Brito, B. Gredler-Grandl, R. Finocchiaro, R.F. Veerkamp, Y. de Haas and F. Miglior, 2019. Invited review: Determination of large-scale individual dry matter intake phenotypes in dairy cattle. J. Dairy Sci. 102: 7655–7663.

Modelling metabolic efficiency – Do we need to understand the biological meaning of residual feed intake breeding values?

M.H. Lidauer[1], T. Andersen[2], U.S. Nielsen[2], E.A. Mäntysaari[1] and G.P. Aamand[3]

[1]Natural Resources Institute Finland (Luke),Genomics and breeding, Tietotie 4, FI-31600 Jokioinen, Finland
Corresponding Author: martin.lidauer@luke.fi
[2]SEGES Innovation P/S, Agro Food Park 15, DK-8200 Aarhus N, Denmark
[3]Nordic Cattle Genetic Evaluation, Agro Food Park 15, DK-8200 Aarhus N, Denmark

Abstract

Selecting dairy cattle that are efficient in utilizing feed is topical for many reasons discussed widely. Residual feed intake (RFI) is a trait that can capture the metabolic efficiency of a dairy cow in a broad sense including the abilities to digest feed, to have low energy loss through methane exhalation, and to use metabolizable energy for production. In 2020 the countries Denmark, Finland and Sweden have included the traits metabolic body weight and RFI into the Nordic Total Merit indices for Holstein, Nordic Red and Jersey dairy cattle. When adding more data from commercial farms to the genomic prediction for RFI, it was recognized that partial regression coefficient estimates may vary significantly across environments and parities. In this study we applied different approaches to predict breeding values for metabolic efficiency in Jersey dairy cows to assess the implications of using RFI breeding values for selection.

The approaches were:

A) similar to the current Nordic RFI evaluation where firstly dry matter intake (DMI) is regressed on energy sinks to get RFI observations that are subsequently used for predicting breeding values.

B) was otherwise same as A) but instead of regressing DMI on energy sinks, DMI was regressed on expected DMI, where expected DMI values were calculated by utilizing energy requirement estimates from dairy nutrition studies.

C) also otherwise same as A) but RFI observations were the difference between DMI and expected DMI.

D) a random regression breeding value prediction model where DMI was regressed on expected DMI (ReFI).

Estimated heritability was 0.14, 0.16, 0.10 and 0.10, and estimated additive genetic standard deviation was 0.89 kg, 1.02 kg, 0.95 kg and 5.4 % when applying approach A, B, C and D, respectively. Metabolic efficiency was unfavourable correlated with yield traits when applying A or B. In contrast, metabolic efficiency was favourable correlated with yield traits when applying C or D. We found that with approach A and B it was not possible to model the expected feed intake properly, which caused the discrepancy between the breeding values. Consequently, when selecting the genetically 10% best cows based on approach A, then these cows had only a 4% higher feed conversion efficiency compared to average cows, but when selection was based on approach D, then the cows had highest yield and 12 % higher feed conversion efficiency compared to average cows. Results indicate that using RFI as a trait to improve metabolic efficiency in dairy cows should be reconsidered if the modelling of partial regression coefficients for the energy sinks is not satisfying.

Keywords: Feed efficiency, residual feed intake, regression on expected feed intake, Jersey.

Introduction

Improving resources efficiency of dairy cows not only by selecting for the traits commonly included in total merit indices but preferable by also including efficiency traits into total merit indices has been studied intensively. Pryce *et al*. (2015) proposed to include a subindex that accounts for feed saved due to lower feed requirement for maintenance and higher metabolic efficiency of the cow. In this regard, the metabolic efficiency of a cow is understood in a broad sense that includes the abilities to digest feed, to have low energy loss through methane exhalation and to use metabolizable energy for the different energy pathways, and all that can be modelled by residual feed intake (RFI) (Koch *et al*., 1963). For breeding value estimation purposes, it is common that RFI observations are modelled from the same data that is used for estimating the breeding values. Therefore, dry matter intake (DMI) is modelled by partial regression analyses and obtained deviations from the expectations, i.e. the residuals, are used subsequently as observations for estimating RFI breeding values (Berry and Crowley, 2013). Alternatively, RFI can be described by a linear function of feed intake and energy sink traits breeding values from a multivariate analysis (Kennedy *et al*., 1993), which also can be tailored to complex multiple-trait random regression models (Islam *et al*., 2020).

In 2020 Nordic Cattle Genetic Evaluation (NAV) has included a Saved feed index into the Nordic Total Merit for selecting Holstein, Nordic Red and Jersey dairy cattle in Denmark, Finland and Sweden (Stephansen *et al*., 2021). The Saved feed index considers the animal’s genomic breeding value for metabolic body weight (MBW) to account for the feed needed for maintenance, and the animal’s genomic breeding value for RFI to accounts for the metabolic efficiency of the cow. Research on improving genomic predictions for both traits has been continued with the aim to increase the reliabilities of the genomic breeding values. For the MBW evaluation carcass weight will be included as correlated trait and for RFI more data from commercial herds with the CFIT 3D camera system (Lassen *et al*., 2018) are used. In this attempt, we have recognized that partial regression coefficient estimates from the first step of the RFI evaluation may vary significantly across environments and parities. Furthermore, partial regression coefficients also differed from regression coefficients developed in nutrition studies (Agnew *et al*., 2003). This was observed both on research farm data and CFIT data. In an earlier study (Mehtiö *et al*., 2018) it was also found that the partial regression coefficients, when estimated from the data, may deviate significantly from those obtained in nutrition studies. Moreover, we noticed that the varying of partial regression coefficient estimates has a noteworthy effect on RFI breeding values. Therefore, a better understanding of the implications of using RFI breeding values for genetic improvement of dairy cattle is needed.

An alternative approach (Lidauer *et al*., 2022) that allows modelling regression coefficients to be close to their biological expectations, is regression on feed intake (ReFI), where DMI is regressed on expected DMI (eDMI). For this approach, eDMI is calculated by using energy requirement (ER) coefficients that are estimated in nutrition studies (e.g. Agnew *et al*., 2003) rather than estimating partial regression coefficients simultaneously while estimating breeding values for RFI. The aim of this study was to contrast estimated breeding values (EBV) obtained from a RFI model like that applied by NAV with those obtained from the ReFI model. To dissect differences between the two approaches, we modelled metabolic efficiency by two additional approaches and estimated the genetic correlations between these metabolic efficiency traits and the production traits milk yield (MY), fat yield (FY), and protein yield (PY). We also contrasted phenotypes of selected superior cows, when selection is based on the four alternative EBV.

Material and methods

Data

For this study Jersey feed efficiency data were extracted from the NAV genetic evaluation. The data were collected on four herds between January 2019 and December 2021 and included 46,822 records up to the sixth parity of 1,211 Jersey cows (Table 1). Weekly averages of daily DMI and body weight (BW) predictions were attained by the CFIT 3D camera system (Lassen *et al*., 2018). The BW predictions were used to calculate daily MBW and daily BW change (DBW). The milk yield data were from the official milk recording scheme and were used to form for each intake record an energy corrected milk (ECM), MY, FY and PY observation. All data was cleaned from outliers.

Table 1. Descriptive statistics for dry matter intake (DMI), metabolic body weight (MBW), energy corrected milk (ECM), milk yield (MY), fat yield (FY) and protein yield (PY) given for 1,211 Jersey cows by first (1) and later (2+) parities.

Parity	N		DMI	MBW	ECM	MY	FY	PY
1	18,221	Mean	21.3	95.0	28.9	22.0	1.3	1.0
		Std	3.2	5.5	5.6	4.9	0.3	0.2
2+	28,588	Mean	24.5	103.3	37.1	28.5	1.7	1.2
		Std	3.3	5.1	6.1	5.6	0.3	0.2

The RFI approach and ReFI approach are conceptually rather different. Despite that the ReFI approach utilizes coefficients from ER studies, also the applied model is rather different. To better understand the source of differences in EBV from both approaches, we also included two intermediate approaches.

Approach A: Currently used residual feed intake model (RFI_A)

The approach involved two steps. In a first step DMI was regressed on energy sinks to get RFI observations:

$$DMI = c_1 \times A + c_2 \times A^2 + LP + LYS + HYS + \gamma_1 \times ECM + \gamma_2 \times MBW + \gamma_3 \times \Delta BW + rfi \quad (1)$$

where c_1 and c_2 are regression coefficients to model the calving age (A) effect of the cow; LP is lactation month nested within first and later parities; LYS is lactation period nested within year and season; HYS is the herd × year × season contemporary effect; γ_1, γ_2 and γ_3 are the coefficients for the partial regressions on the energy sinks ECM, MBW and ΔBW, respectively that are nested within LP classes; and rfi is the model residual and was regarded as a raw RFI observations. The raw rfi observations were adjusted for heterogeneous variance to accommodate a single trait evaluation in the second step (Stephansen *et al*., 2021). The final RFI observations used in the second step were formed by adding to the adjusted rfi observations the LYS and HYS solutions from model (1) to avoid that genetic variation would be lost in step 1. For the second step the applied breeding value estimation model was:

$$\mathrm{RFI} = \mathrm{LYS} + \mathrm{HYS} + \mathrm{pe} + \mathrm{a} + \mathrm{e} \qquad (2)$$

where LYS and HYS are same fixed effects as described for model (1); pe is the cow's random permanent environmental effect; a is the random additive genetic animal effect; and e is the random residual.

Approach B: Residual feed intake with regression on feed requirement (RFI_B).

The approach was otherwise identically with approach A, but the model for the first step was modified:

$$\mathrm{DMI} = c_1 \times A + c_2 \times A^2 + \mathrm{LP} + \mathrm{LYS} + \mathrm{HYS} + \phi \times \mathrm{eDMI} + \mathrm{rfi}, \qquad (3)$$

where all effects in the model are the same as in model (1) but the model included instead of partial regressions on energy sinks a regression on eDMI, where ϕ is a regression coefficient nested within LP classes. Calculating for each record the eDMI value was done by firstly calculating ER:

$$\mathrm{ER} = 4.81 \times \mathrm{ECM} + 0.603 \times \mathrm{MBW} - 27.6 \times \mathrm{BW\_Loss} + 38.3 \times \mathrm{BW\_Gain} \qquad (5)$$

where the coefficients are metabolizable ER in mega joules for producing 1 kg ECM, maintaining 1 $\mathrm{kg}^{0.75}$ MBW, utilizing energy from 1 kg BW loss, and gaining 1 kg BW. The applied coefficients were the averages of the estimates reported by Agnew *et al*. (2003). Then, ER values were scaled so, that the obtained eDMI covariables should yield ö regression coefficient estimates that have on average an expectation of unity:

$$\mathrm{eDMI} = \mathrm{ER} \times \mathrm{mean(DMI)} / \mathrm{mean(ER)}. \qquad (6)$$

Approach C: Requirement residual feed intake (RFI_C).

Again, the approach was otherwise identically with approach A, but a raw RFI observation was calculated as a difference:

$$\mathrm{rfi} = \mathrm{DMI} - \mathrm{eDMI} \qquad (7)$$

Approach D: Regression on feed intake (ReFI).

This approach only requires modelling of DMI by a random regression model that regresses DMI on eDMI:

$$\text{DMI} = \beta \times \text{eDMI} + \eta \times \text{eDMI} + y \times \text{eDMI} + \acute{a} \times \text{eDMI} + e \qquad (8)$$

where â is a fixed regression coefficient nested within herd × year × parity classes, h is a random regression coefficient nested within herd × year × month classes, is a random regression coefficient nested within permanent environmental classes, á is a random regression coefficient nested within additive genetic animal classes, and e is the random residual.

Estimation of variance components and breeding values

The relationships between animals were modelled by setting up the numerator relationship matrix (**A**). Therefore, the pedigree of all cows with observations was pruned including five generations at most, which resulted 4,223 informative animals in the pedigree. Univariate analyses were carried out to estimate variance components for the four metabolic efficiency approaches by applying model (2) for the RFI approaches (A, B and C) and model (8) for the ReFI approach (D). The genetic correlations between the four metabolic efficiency traits and the production traits MY, FY and PY were estimated by multivariate analyses using as observations yield deviations (YD) and applying a multiple-trait repeatability animal model that included a mean and the random permanent environmental, random animal and random residual effects. The YD observations were obtained by firstly carrying out a multivariate analysis for the yield traits and univariate analyses for metabolic efficiency by the approaches A, B, C and D, and followed by summing solutions for the permanent environmental, additive genetic, and residual effects. For the genetic correlation study only first parity data was used.

All data were used for the prediction of EBV based on model (2) for the RFI approaches (A, B and C) and model (8) for the ReFI approach (D). Cows with at least 5 observations in the first parity were ranked alternatively based on the four different sets of EBV, and the first parity phenotypic means of the genetically 10 % best cows were contrasted against the phenotypic means of all cows.

Results and discussion

Variance component estimates

The estimated heritability was 0.14, 0.16, 0.10 and 0.10, and the estimated additive genetic standard deviation was 0.89 kg, 1.02 kg, 0.95 kg and 5.4 % for RFI approach A, B, C and ReFI approach, respectively. Considering that the average DMI was 23.2 kg, the genetic standard deviations estimated by approach A, B and C can be expressed in efficiency percentages to make it comparable to ReFI. Thus, the genetic standard deviations for RFI by approach A, B, and C relate to 3.8 %, 4.1 % and 5.2 %, respectively, which all are lower than that one estimated by ReFI. Even estimated genetic variance was higher for RFI_C and ReFI, the heritability was lower. This was because RFI_C and ReFI resulted higher residual variance estimates compared to those from RFI_A and RFI_B.

The genetic correlation between metabolic efficiency by the four approaches and yield traits are given in Table 2. For RFI_A and RFI_B we obtained moderate positive (unfavourable) correlations with the yield traits. This was unexpected, because RFI observations are corrected for yield. In contrast, for RFI_C and ReFI we obtained

Table 2. Genetic correlations between metabolic efficiency by approach A (RFI_A), B (RFI_B), C (RFI_C) and D (ReFI), and the yield traits milk (MY), fat yield (FY) and protein yield (PY).

Yield trait	RFI_A	RFI_B	RFI_C	ReFI
MY	0.47	0.43	-0.01	0.02
FY	0.17	0.24	-0.20	-0.28
PY	0.31	0.30	-0.16	-0.10

moderate negative correlations with FY and PY, which indicates a favourable genetic association between metabolic efficiency and milk content traits, in particular with fat %.

Regression coefficients for modelling feed intake expectations

The differences in the estimated genetic correlations between the yield traits and metabolic efficiency by the four approaches are a result on how the expected DMI is modelled. For RFI_A the expectations were obtained by the estimated partial regression coefficients for the three energy sink traits ECM, MBW and DBW (Table 3). For RFI_B only one single regression coefficient (ö) was fit for each class, instead of fitting a partial regression coefficient triplet. For approach RFI_C no regression coefficients were estimated, and for the ReFI approach regression coefficients (â) were estimated within a different set of classes because regression coefficients also modelled the contemporary group effect.

The mean estimates given in Table 3 indicate that there is a discrepancy between the estimated partial regression coefficients and the coefficients reported in ER studies (e.g. Agnew *et al*., 2003). Considering that ER regression coefficients applied in equation (5) are sufficiently close to the ER of the cows in this study, and that the average energy density of the feed was 9.80 mega joule / kg DMI, then it can be expected that on average a cow would have required 0.49 kg, 0.062 kg, and 3.91 kg DMI to produce 1 kg ECM, maintain 1 $\text{kg}^{0.75}$ MBW and increase BW by 1 kg, respectively. However, the partial regression coefficient estimates for ECM were on average 37% and 63% lower in first and later parity, respectively. And the partial regression coefficients for MBW were 2.6 and 3.2 times larger in first and later parity, respectively. Consequently, for high yielding cows the modelled feed intake expectations were far too low and such cows received a positive RFI observation, which explains the unfavourable correlation between RFI_A and yield traits. Also the regression coefficients for RFI_B where significantly lower than the expected value of 1.0. Therefore, similarly to RFI_A, for high producing cows the modelled feed intake expectations were too low, which

Table 3. Means of regression coefficient estimates across lactation month classes (RFI_A, RFI_B) and across herd × year classes (ReFI) by first (1) and later (2+) parities. Energy sink traits: energy corrected milk (ECM), metabolic body weight (MBW), body weight change (∆BW), and expected dry matter intake (eDMI).

	RFI_A			RFI_B	ReFI
	ECM	MBW	∆BW	eDMI	eDMI
Parity	γ_1	γ_2	γ_3	φ	β
1	0.307	0.159	-0.277	0.463	1.036
2+	0.179	0.196	-0.244	0.322	1.013

resulted also in unfavourable correlations with the yield traits. For the ReFI approach the estimated regression coefficients were on average close to the expected value of 1.0 (Table 3), which supports that the estimated favourable correlations with yield traits might be closer to the true genetic association between metabolic efficiency and yield. The finding for the ReFI approach were supported by the RFI_C approach, which did not require to estimate regression coefficients from the data.

Phenotypes of genetically superior cows based on four alternative breeding values

The four different sets of EBV for metabolic efficiency differed considerably. We obtained a correlation between EBV from the approach A with EBV from approaches B, C and D of 0.92, 0.69 and 0.61, respectively. Consequently, groups of genetically superior cows were rather different when selecting based on the different sets of EBV. In Table 4 the first parity phenotypic means of the four different cow groups are tabulated against the first parity phenotypic means of all cows with at least five first parity observations.

When 10 % best cows were selected based on RFI_A and RFI_B then they had lowest DMI but also lower yields compared to average cows. In contrast, when 10 % best cows were selected based on ReFI, then cows had lower DMI as average cows but highest yields. Consequently, when selecting based on RFI_A, then the feed conversion efficiency of the 10 % best cows was only 4 % higher than that one of average cows, but the feed conversion efficiency was 12% higher when selection was based on ReFI.

Conclusion

In this study we compared different approaches to predict breeding values for metabolic efficiency in dairy cows. With the currently used approach that models metabolic efficiency by a residual feed intake model, genetically superior cows were only moderately better in feed efficiency but had lower production. In contrast, when applying an alternative approach, where dry matter intake is regressed on expected dry matter intake, then superior cows were clearly more efficient and had also significantly higher production. The poor performance of the residual feed intake model in this study was due to the inability of the model to estimate the partial regression coefficients for the energy sinks properly.

Table 4. Phenotypic means of first parity dry matter intake (DMI), metabolic body weight (MBW), energy corrected milk (ECM), milk yield (MY), fat yield (FY), protein yield (PY) and feed conversion efficiency (FCE=ECM/DMI) for all cows with at least 5 observations and for 10% best cows with at least 5 observations selected based on EBV estimated by four Approaches. Approach A (RFI_A), B (RFI_B), C (RFI_C) and D (ReFI).

All cows		**DMI**	**MBW**	**ECM**	**MY**	**FY**	**PY**	**FCE**
10% best cow	**Method**	**21.2**	**94.9**	**28.7**	**21.8**	**1.30**	**0.946**	**1.35**
	RFI_A	18.6	94.7	26.3	18.8	1.22	0.860	1.41
	RFI_B	18.4	92.8	25.3	18.3	1.17	0.833	1.38
	RFI3_C	19.9	93.1	29.1	21.5	1.34	0.966	1.46
	ReFI	19.8	93.9	29.8	22.0	1.37	0.982	1.51

Acknowledgement

The authors wish to acknowledge funding received from the Innovation Fund Denmark, the Finnish Ministry of Agriculture (453/03.01.02/2018) and the dairy farmer's owned companies and corporations VikingGenetics Ltd Denmark, Valio Ltd Finland, Nordic Cattle Genetic Evaluation NAV Denmark, and Faba co-op Finland.

List of References

Agnew, R.E., T. Yan, J.J. Murphy, C.P. Ferris and F.J. Gordon. 2003. Development of maintenance energy requirement and energetic efficiency for lactation from production data of dairy cows. Livestock Production Science 82, 151-161.

Berry, D. and J. Crowley. 2013. Cell biology symposium: genetics of feed efficiency in dairy and beef cattle. Journal of Animal Science 91(4):1594-1613.

Islam, M.S., J. Jensen, P. Løvendahl, P. Karlskov-Mortensen, and M. Shirali. 2020. Bayesian estimation of genetic variance and response to selection on linear or ratio traits of feed efficiency in dairy cattle. Journal of Dairy Science 103, 9150-9166.

Kennedy, B. W., J.H.J. Vanderwerf and T.H.E Meuwissen. 1993. Genetic and statistical properties of residual feed-intake. Journal of Animal Science 71, 3239–3250.

Koch, R.M., L.A. Swiger, D. Chambers and K.E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486–494.

Lassen, J., J. R. Thomasen, R. H. Hansen, G. G. B. Nielsen, E. Olsen, P. R. B. Stentebjerg, N. W. Hansen and S. Borchersen. 2018. Individual measure of feed intake on in-house commercial dairy cattle using 3D camera system. Page 2018 in Proc. Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand.

Lidauer, M.H., E. Negussie, T. Mehtiö, P. Mäntysaari, T. Kokkonen and E.A. Mäntysaari. 2022. Modelling multiplicativity in feed efficiency by regression on expected feed intake. Proceedings, 12th World Congress of Genetics Applied to Livestock Production, Rotterdam, The Netherlands, July 3-8, 2022, accepted, [4 p.]

Mehtiö, T., E. Negussie, P. Mäntysaari, E.A. Mäntysaari and M.H. Lidauer. 2018. Genetic background in partitioning of metabolizable energy efficiency in dairy cows. J. Dairy Sci. 101:4268–4278.

Pryce, J. E., O. Gonzalez-Recio, G. Nieuwhof, W. J. Wales, M. P. Coffey, B. J. Hayes and M. E. Goddard. 2015. Definition and implementation of a breeding value for feed efficiency in dairy cows. J. Dairy Sci. 98:7340–7350.

Stephansen, R.S., M.H. Lidauer, U.S. Nielsen, J. Pösö, F. Fikse, C.I.V. Manzanilla-Pech and G.P. Aamand. 2021. Genomic prediction of residual feed intake in the Nordic breeds using data from research herds and 3D cameras in commercial herds. In: Proceedings of the 2021 Interbull meeting, Leeuwarden, The Netherlands, April 26-30, 2021. Interbull Bulletin 56:162–166.

Large scale phenotyping of methane for genetic evaluation is possible with ‘Sniffers’

M.N. Aldridge[1], A. van Breukelen[1], R.F. Veerkamp[1] and Y. de Haas[1]

[1]Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands
Corresponding Author: michael.aldridge@wur.nl

Abstract

Our objective was to demonstrate that the ongoing recording of methane of individual cows on 100 dairy farms in the Netherlands will provide meaningful breeding values for methane mitigation. Previously we conducted a power calculation that used the available literature, which predicted that recording on 100 dairy farms, 15,000 cows, and two years of recording, would provide a high enough reliability of prediction, to allow for practical direct selection for lower methane emissions. To determine what reduction in methane could be achieved by 2050, a methane trait with genomic prediction was included in the Dutch national selection index, showing that methane intensity could be reduced by between 14% and 24% (depending on the strategy). To be able to achieve these gains, a large scale phenotyping strategy needed to be implemented. We have commissioned 100 infrared spectrometry methane sensors, or ‘sniffers’, to be used to fulfil the data need. To date, 15 sensors have been installed, and at the time of writing, installation of the remaining sensors is ongoing and will be completed in the second half of 2022. The benefit of using sniffers is that it allows for a high throughput of cows and continuous recording for extended periods of time. With the data already available we have calculated traits referred to as mean visit, daily, and weekly methane concentration (ppm). These methane concentration traits are moderately heritable (0.13 to 0.32) and repeatable (0.30 to 0.68). The downside of using sniffers is that it records the methane concentration and not the methane production (g/day). We have also temporarily installed GreenFeeds on 16 farms (four overlapping with sniffers). There are 1,800 cows with sniffer recorded weekly methane concentration, and 822 cows with GreenFeed recorded weekly methane production (184 cows with both). All of the cows are linked via the pedigree and there are 1,655 which are also genotyped. A preliminary genetic correlation between Sniffer methane concentration and GreenFeed methane production is high at 0.71, indicating that selection for methane concentration as recorded with sniffers will in fact reduce methane emissions. We have learned a number of valuable lessons in the large scale rollout of methane sniffers which will be useful for future industry application. Most importantly, is that suitably reliable genomic breeding values for methane are closer to being a reality than ever before.

Keywords: Greenhouse gas, environment, dairy cows, ruminant, CH_4, CO_2

Introduction

The Dutch government has committed to reducing greenhouse gas emissions by 55% in 2030 compared to year 1990, and by 2050 to be carbon neutral (Communication from the Directorate-General for Climate and Energy, 2022). The Dutch dairy industry is a large contributor to greenhouse gases in the Netherlands (Van der Maas *et al*., 2009). We have been investigating how to use animal breeding as a methane mitigation tool, to help achieve these goals and reduce the environmental impact of the Dutch dairy industry.

To achieve a reduction in methane with breeding, accurate breeding values are required. To obtain accurate breeding values large scale recording of methane is needed. The use of various infrared spectrometry sensors have been used in various studies to estimate genetic parameters of methane (Lassen and Difford, 2020). Collectively these sensors have been referred to as 'sniffers'. Sniffers offer a comparatively cheap method of phenotyping a large number of commercial cows (Madsen *et al.*, 2010; Garnsworthy *et al*., 2019).

We have reviewed our decision process in selecting sniffers as the most suitable method of large scale recording and how we determined that 100 farms would need to have a sniffer installation, to obtain reliable genomic breeding values (de Haas *et al.,* 2021). We summarise the parameter estimates obtained from large scale methane recording with sniffers, including the trait definitions and genetic correlations with methane traits measured with GreenFeed (C-lock Inc. Rapid City, SD, US (Zimmerman, 2011)). Then using the current knowledge and lessons learned during the upscaling to 100 farms, we outline the next steps towards genomic breeding values and implementation in breeding programs. This demonstrates the progress made towards accomplishing our objective, of large scale phenotyping with sniffers in the Netherlands to provide meaningful breeding values for methane mitigation.

Materials and methods

Current projects

Wageningen University and Research has ongoing research on the application of breeding as a strategy for methane mitigation. The results presented here are summaries of the work done during the projects Climate Envelope (The Klimaatenvelop is supported by the Dutch Ministry of Agriculture, Nature and Food quality), and Climate Smart Cattle Breeding (Public-private partnership with the cooperative cattle improvement organization CRV, Friesland Campina, and the Dutch government). We have focused only on recording and outcomes of ongoing projects directly related to animal breeding.

Recording methane on commercial farms

Determining the number of farms to record with sniffers

We wanted to determine what the current genetic trends are for methane production (g/day) and methane intensity (g/kg), and what theoretical gain could be made if methane was added to the Dutch national breeding goal. To this end, we adapted a selection index that included 15 traits that are included in the current breeding goal. We tested various goals and economic weights placed on methane with or without genomic prediction. Concurrently a power analysis was performed to determine a recording strategy that would achieve a high enough reliability of genomic prediction that would facilitate the selection and genetic gain seen in the selection index. Published genetic parameters and preliminary data collected during the Climate Envelope were used as inputs for the selection index and power analysis. For full details on the selection index and power analysis see de Haas *et al*. (2021).

Developing a sniffer and installation

Previously 15 sniffers were purchased and installed on 14 commercial dairy farms across the Netherlands (WD-WUR v1.0, manufactured by Carltech BV), between March 2019 and September 2020. The sniffers were installed with the air intake located in the feeding bin of automated milking robots (AMS), where they measured methane concentration (0 to 2,000 ppm) and carbon dioxide concentration (0 to 10,000 ppm) continuously, for between 64 and 436 days (depending on farm). During the data collection and analysis, several practical and technical lessons were learnt. These lessons were used to develop a second generation of sniffers. Installation of an additional 90 units has begun using the new design. Methane data has continued to be collected after September 2020 and is expanding as new sensors are installed. The data collected from the sniffers is uploaded in near real time to a Microsoft Azure data cloud.

Defining methane traits from sniffer recorded methane

Using long-term and continuous recording with sniffers can be used to defined methane traits, but the raw sniffer requires processing for this purpose. A method similar to Garnsworthy *et al*. (2012) was used to match cow visits to the AMS with peaks in methane concentration. The data was then processed to remove background levels of methane and carbon dioxide and correct for cow behaviour. Across the farms there were 1,746 Holstein Friesian cows that visited an AMS a total of 308,968 times, during each milking event the methane concentration was measured.

The main considerations included, calculation methods (mean, median, log, ratio) and the period used per record (visit or weekly). Based on various combinations of trait definitions we wanted to estimate heritability and repeatability of various methane traits defined from the largest data set of long-term repeated cow records. All cows were linked via pedigree and 1,611 were genotyped. Parameter estimates were made using univariate animal models, that included repeated records, with ASReml 4.2 (Gilmour *et al*., 2015). The heritabilities and repeatabilities were used to calculate reliability of the breeding values and determine the number of records per daughter that would be needed for implementation. For full details on matching AMS and sniffer data, data processing, parameter estimations, and determining reliability of estimated breeding values see van Breukelen *et al*. (2022a).

Sniffer and GreenFeed methane traits

The application of sniffers for genetic selection is still in a development stage, and is why we explored the genetic relationship with traits recorded with GreenFeed, as GreenFeed are considered the gold standard for on-farm recording of individual cow methane. Since the univariate analysis with the sniffers, data collection has been ongoing and is expanding to measure methane on 100 dairy farms. There have also been GreenFeed units installed temporarily on 16 farms (four overlapping with sniffers). At the time of this analysis, there were 1,800 cows with sniffer recorded daily methane concentration, and 822 cows with GreenFeed recorded daily methane production (g/day) (184 cows with both). All of the cows are linked via the pedigree and there are 1,655 which are also genotyped. Bivariate animal models in ASReml 4.2 (Gilmour *et al*., 2015) were used to estimate the genetic parameters for methane concentration (ppm recorded by sniffers) and methane production (g/day recorded by GreenFeed).

Results and discussion

Why and how to record methane on commercial farms?

Based on the current Dutch selection index, the genetic trend for methane production is unfavourable with a genetic trend of increasing methane by 5.79 g/day per generation. However, as the current index selects for traits that improve production efficiency, methane intensity will decrease by 13%. With active selection on lowering methane production and utilizing genomic prediction, the genetic trend for methane production can be decreased by 5.79 g/day, will still result in increasing milk production, and lower methane intensity by 24%. This was a clear justification for us to develop a recording scheme that will allow us achieve these theoretical improvements. We found, to achieve a desired reliability of genomic prediction of 0.40 within two years of recording, we would need to record methane on 100 farms with an average herd size of 150 cows (15,000 total cows). For full results of the selection index and power analysis see de Haas *et al*., (2021).

During the data collection and analysis of the first 15 sniffers, we identified several technical requirements that should be considered when developing new sniffers. Due to the challenging environment of dairy barns and the sensitivity of methane sensors, the housing containing the sensor should be dust resistant, operate within a wide range of ambient temperatures and humidity, and exclude gases that can interfere within the measured wavelength such as water. Methane sensors can have a large range of detection limits, while most measurements are within 200 to 2,000 ppm, consider that the error should be minimized within the most common recording ranges (measurements closer to the mean per visit should be the most accurate). Some methane sensors are prone to drifting away from the calibration curve, limiting factors that affect the methane sensor can help to minimize the drift, regular checks of the incoming data can help identify when a sensor drifts and requires recalibration. If this is ignored it can limit the potential analysis, for example traits that use a ratio with carbon dioxide were limited because the methane sensor drifted and the carbon dioxide sensor did not drift. The methane data recorded in these projects is uploaded in near real time to a data cloud infrastructure, however there can be issues with communication due to interference with farm equipment or with poor connections via the telecommunication infrastructure, having built-in data storage options are a good redundancy. With these lessons learned, 90 new sniffers were purchased that implement these requirements. However, they are still in development and with limitations in global supply chains the full scale installation was inhibited. The installation on 100 farms is expected to be completed in 2022, in the meantime this still remains the largest and longest methane recording dataset on individual cows for genetic evaluations.

Methane traits can be measured with sniffers and used in genetic evaluations

We have found that mean methane per visit is heritable (0.13 ± 0.01) and repeatable (0.30 ± 0.01), but with lower estimates compared to weekly methane (0.32 ± 0.01 and 0.68 ± 0.01, respectively). However this has limited effect for the reliabilities of the breeding values. To achieve a reliability of 0.50, 25 mean methane records on 10 daughters is needed. For the full results of on the genetic parameters estimated with sniffers see van Breukelen *et al*. (2022a). The estimated heritabilities and repeatabilities, were within the parameter space of the estimates used in the selection index and power analysis, indicating the large scale methane recording currently in progress, will make the goal of providing meaningful breeding values for methane mitigation possible.

The previous parameter estimates were for methane concentration traits, while the desired trait change is placed on methane production. The preliminary results with the GreenFeed have been very encouraging. The heritability (0.20) and repeatability (0.34) for methane production are within the same parameter space as the sniffers (van Breukelen *et al*. 2022b, accepted). Importantly the genetic correlation between GreenFeed and sniffer methane is high and positive (0.71). This indicates that selection for lower methane concentration will reduce methane production and be a useful methane mitigation option. This has given us confidence that by the end of the project there will be the required data for reliable breeding values for a methane trait that can be selected.

Genomic breeding values and implementation in breeding programs

The work so far has mostly utilized data already collected. As the dataset from the full 100 farms becomes available there are several research objectives to be completed. Genetic and phenotypic correlations will be estimated with all of the traits included in the current national Dutch breeding goal. Subsequently a selection index that incorporates the new methane traits will be developed. Reliable breeding values for the methane trait to be included in the selection index need to be estimated. In the meantime, the full 100 sniffers need to be installed, the data processing is being refined, and methods to improve the accuracy of the methane traits are being explored such as standardising the methane measurements across devices and including the microbiome population as a random effect.

Conclusion

Animal breeding is a promising method of methane mitigation that can be used to help reach greenhouse gas targets. Before breeding strategies can be implemented, there needs to be enough data to reliably predict breeding values. Long-term measuring with sniffers is a viable method of data collection. Results so far are encouraging and the traits developed could be used for selection purposes. With the expansion of sniffers to 100 farms we believe meaningful breeding values for methane mitigation will be available in the relatively near future.

Acknowledgement

The research leading to these results was part of the Climate Approach Project, funded by the Dutch Ministry for Agriculture, Nature and Food Quality (LNV, The Hague; project number BO-53-003). This work was supported by the TKI Agri and Food project LWV19155 (Wageningen, the Netherlands) and the partners CRV (Arnhem, the Netherlands) and FrieslandCampina (Amersfoort, the Netherlands). The authors gratefully acknowledge the farmers for participating in this research.

References

de Haas, Y., R.F. Veerkamp, G. de Jong and M.N. Aldridge, 2021. Selective breeding as a mitigation tool for methane emissions from dairy cattle. Animal 15:100294.

Garnsworthy, P.C., J. Craigon, J.H. Hernandez-Medrano and N. Saunders, 2012. On-farm methane measurements during milking correlate with total methane production by individual dairy cows. J. Dairy Sci. 95:3166–318

Garnsworthy, P.C., G.F. Difford, M.J. Bell, A.R. Bayat, P. Huhtanen, B. Kuhla, J. Lassen, N. Peiren, M. Pszczola, D. Sorg, M.H.P.W. Visker, and T. Yan, 2019. Comparison of methods to measure methane for use in genetic evaluation of dairy cattle. Animals (Basel) 9:837

Gilmour, A., B. Gogel, B. Cullis, S. Welham, and R. Thompson, 2015. ASReml User Guide: Release 4.1: Structural Sp ecification. VSN International Ltd.

Lassen, J. and Difford, G.F, 2020. Genetic and genomic selection as a methane mitigation strategy in dairy cattle. Animal, 14:s473-s483.

Madsen, J., B.S. Bjerg, T. Hvelplund, M.R. Weisbjerg, and P. Lund, 2010. Methane and carbon dioxide ratio in excreted air for quantification of the methane production from ruminants. Livest. Sci. 129: 223–227.

van Breukelen, A.E., M.A. Aldridge, R.F. Veerkamp and Y. de Haas, 2022a. Genetic parameters for repeatedly recorded enteric methane concentrations of dairy cows. J. Dairy Sci. 105:4256–4271

van Breukelen, A.E., M.N. Aldridge, R.F. Veerkamp, L. Koning, L.B. Sebek and Y. de Haas, 2022b. Heritability and genetic correlations for enteric methane emissions of dairy cows measured by sniffers and GreenFeed. Proceedings of the 12th World Congress on Genetics Applied to Livestock Production.

Van der Maas, C.W.M., L.J. Brandes, K. Baas, G.J. Van den Born, G. Geilenkirchen, R. Te Molder, D.S. Nijdam, J.G.J. Olivier, C.J. Peek and M.W. Van Schijndel, 2009. Greenhouse Gas Emissions in the Netherlands 1990-2007. National Inventory Report 2009. Netherlands Environmental Assessment Agency, Bilthoven, the Netherlands.

Zimmerman, P, 2011. USA Patent No. US Pat. No. 7966971B2.

A pipeline for the integration of growth, feed efficiency and greenhouse-gas emission data in Italian Holstein

F. Tiezzi[1], R. Finocchiaro[2], L. Benzoni[2], G. Invernizzi[3], J.B.C.H.M. van Kaam[2], M. Marusi[2], G. Savoini[3], R. Bozzi[1] and M. Cassandro[2,4]

[1]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Piazzale delle Cascine 18, 50144 Firenze (FI), Italy
Corresponding Author: francesco.tiezzi2@unifi.it
[2]Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana (ANAFIBJ), Via Bergamo 192, 26100 Cremona (CR), Italy
[3]Dipartimento di Medicina Veterinaria e Alimentazione Animale, Via dell'Università 6, 26900 Lodi (LO), Italy;
[4]Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD),Italy

Abstract

The objective of this study was to describe the collection protocol for growth, feed efficiency and greenhouse gas (GHG) emissions in young Italian Holstein bulls. The phenotypes will be used in genetic evaluations for the reduction of environmental impact in Italian Holstein.

The animals involved in this study were young Italian Holstein bulls undergoing own performance test at the Genetic Center (GC) of Italian National Breeders Association for the Holstein, Brown and Jersey dairy cattle breeds (ANAFIBJ). Phenotypic data was provided for on 218 Holstein bulls between the age of 171 and 541 days. All bulls were genotyped using various SNP chips resulting in 69,127 SNP after imputation.

The phenotypic data can be summarized in three groups:

1. a group of traits describing growth and condition of the animal;
2. a dataset derived from measures taken with the Roughage Intake Control system (RIC; Hokofarm Group, Marknesse, The Netherlands) and
3. dataset derived from measures taken using the GreenFeed (C Lock Inc., Rapid City, SD, USA).

The group **A** included measures of body growth taken using electronic scales and stadiometers operated by qualified personnel. These included body weight (WEI), body condition score (BCS), heart girth (HG) and height (HEI). Group **B** included single-visit measures that were converted into daily measures and included: number of visits at the feeder per day (NVF), average intake at the feeder (AIF), average time at the feeder (ATF). Group **C** included single-visit measures then converted to daily records and included: number of visits (NVG), carbon-dioxide daily emission (CO_2), methane daily emission (CH_4), average airflow (AIR) and average time (ATG).

Variance components and genetic parameter estimation (heritability and genetic correlations) was carried out using a GBLUP mixed model that included a genomic relationship matrix built on the SNP markers.

The growth traits showed the largest estimates of heritability, close to 0.40. Heritability estimates for the RIC-derived traits were lower, ranging from 0.167 (AIF) to NVF (0.306). Estimates for emission traits ranged from 0.241 for ATG to 0.480 for CO_2.

Results suggest that selection indices could be built in order to reduce GHG emissions without compromising growth, condition, stature and feed intake. The upcoming research should be focused on the use of feed efficiency and relative GHG emission, as these components need to be adjusted by the growth and size of the animal. Subsequently a further data-set should include GC sire performance and their daughters', reared in dairy commercial farms.

Keywords: Greenhouse gas emissions, feed efficiency, genomic selection, Italian Holstein.

Introduction

Dairy cattle is known to be impactful on greenhouse gasses (GHG) emissions, with its enteric emissions accounting for over ten percent of the emissions from the livestock sector globally (Gerber *et al*., 2013). Methane emissions are also energetically expensive because of the fermentation process (Appuhamy *et al*., 2016) but this inefficiency could be reduced by steering the fermentation process, redirecting the energetic resources to reproduction and milk production (Haque, 2018). Methane and carbon dioxide emissions have been shown to be heritable, providing the basis for applying genetic selection for their reduction (Cassandro *et al*, 2010; Cassandro M., 2013; Cassandro *et al*., 2013; Pickering *et al*., 2015; Lassen and Løvendahl, 2016). Such selection could be applied by selecting directly for breath measurements, but also using indirect selection *including* indicator traits such as feed intake (de Haas *et al*., 2017; Niero *et al*., 2020).

The objective of this study was to estimate genetic parameters for verifying the feasibility of (direct or indirect) selection for reduced GHG emissions in Italian Holsteins.

Material and methods

Animals and data

The animals involved in this study were young Italian Holstein bulls undergoing progeny test

in the genetic centre of the Italian National Breeders Association for the Holstein, Brown and Jersey dairy cattle breeds (ANAFIBJ) as reported by Callegaro *et al*. (2022). The ANAFIBJ genetic center is equipped with five Roughage Intake Control system units (RIC; Hokofarm Group, Voorsterweg, The Netherlands) distributed over three pens. One of the three pens is also equipped with the Automated Head-Chamber System (AHCS; GreenFeed C Lock Inc., Rapid City, SD, USA), an automated feeding station designed to measure daily CH_4 and CO_2 emissions (g/d) from ruminant's breath (Hristov *et al*., 2015).

Phenotypic data was made available on 221 Holstein bulls between the age of 171 and 541 days. All bulls were genotyped using various SNP chips resulting in 69,127 SNP. Genomic data was edited using the preGSf90 software (Aguilar *et al*., 2010), removing SNP with call rate below 0.90 and minor allele frequency below 0.05. After editing, 61,591 SNP were available.

The first set of traits included measures of body growth taken using electronic scales and stadiometers operated by qualified personnel. These included body weight (WEI), body condition score (BCS), heart girth (HG) and height (HEI). The second group of traits included measures of feed intake and feeding behaviour assessed using the Roughage Intake Control system (RIC; Hokofarm Group, Marknesse, The Netherlands). Records were organized as daily measures: number of visits at the feeder per day (NVF), average intake at the feeder (AIF), average time at the feeder (ATF). The third group of traits included measures of greenhouse gasses emission and were obtained using the GreenFeed (C Lock Inc., Rapid City, SD, USA). Again, the single-visit measures were converted to daily records and included: number of visits (NVG), carbon-dioxide daily emission (CO_2), methane daily emission (CH_4), average airflow (AIR) and average time (ATG). A single dataset was created, containing all the phenotypic information. Individuals showed an average of 4 records for the growth traits, 36 records for the RIC-derived traits, 19 records for the GreenFeed-derived traits. For all records, the age at recording was calculated. In addition, the date of birth of the individual was transformed to numerical values as the difference, in days, from a fixed arbitrary date.

Statistical analysis

Estimates of variance components and genetic parameters were carried out using a linear mixed model that included a genomic relationship matrix constructed using the SNP markers. The model was defined as follows:

$$y = Xb + Z_d d + Z_p p + Z_a a + e \qquad (1)$$

where **y** is the vector containing phenotypic records, **X** and **b** are the incidence matrix and vector of solutions for the fixed effects (age at phenotyping, date of birth), $\mathbf{Z_d}$ and **d** are the incidence matrix and vector of solutions for the 'date of recording' uncorrelated random effect, $\mathbf{Z_p}$ and **p** are the incidence matrix and vector of solutions for the animal permanent environmental uncorrelated random effect, $\mathbf{Z_a}$ and **a** are the incidence matrix and vector of solutions for the animal additive genetic random effect (with genomic relationship matrix), **e** is the vector of random residuals. The model was implemented for single-trait and two-trait analyses in order to obtain estimates of heritability and genetic correlations, using the gibbs2f90 software (Misztal *et al*., 2002). Estimates of variance components and genetic parameters were obtained as posterior means and their dispersion was obtained as the posterior SD and the 95% confidence intervals. The heritability was expressed as the ratio between the additive genetic variance and the sum of the four variance components.

Results

Descriptive statistics and heritability estimates for the studied traits are reported in table 1. The growth and condition traits showed the largest estimates, all being above or close to 0.40. While these traits are expected to be highly heritable, the estimates appear larger compared to those found in literature. This could be due to the relatively small sample size. On the other hand, heritability estimates for the RIC-derived traits were lower, ranging from 0.167 (AIF) to NVF (0.306). Heritability estimates for the emission traits were moderate to high, ranging from 0.241 for ATG to 0.480 for CO_2.

Genetic correlation estimates are reported in table 2. Among the growth traits, correlations were strong for WEI-BCS and BCS-HG, but weak for all the other combinations of traits. Correlations between NVF and AIF was 0.74, the genetic correlation between AIF and ATF was 0.98 (result not shown in table). Correlations among the GreenFeed-derived traits were moderate to strong, especially among CH_4, CO_2 and AIR, which were all above 0.70. NVG was moderately related to the other GreenFeed-derived traits, with correlations below 0.7. The genetic correlation between AIR and ATG was 0.95 (result not shown in table). Genetic correlations between the growth traits and the RIC-derived traits were all positive and strong, approaching unity

(0.89 to 0.99). Genetic correlations between the growth traits and the GreenFeed-derived traits were also positive and strong, with values ranging from 0.90 to 0.96. Similarly, the genetic correlations between the GreenFeed-derived traits and the RIC-derived traits were positive and moderate to strong, with values from 0.60 to 0.73.

Discussion

Heading of subsection

The estimates show substantial genetic variation for all the studied traits. The CO_2 and CH_4 daily emissions show high heritability with the possibility of selection, therefore reduction of GHG emissions. The estimated values for the heritability of CH_4 and CO_2 are larger than values found in literature (Lassen and Løvendahl, 2016; Brieder *et al*., 2018), although this could be due to the involvement of growing bulls rather than lactating cows and the limited sample size in the current study. The genetic correlations indicate that there is the possibility to select for less impactful animals. The genetic correlation between CO_2 and CH_4 emissions is strong (0.84), suggesting that selection for one component of the emissions would improve the other as well. Similarly, the AIR shows strong correlations with the GHG emissions, suggesting that this trait could be used as an indicator. Similarly, the NVG could serve as an indicator trait to reduce the emissions, although the achievable genetic progress would be limited due to the moderate genetic correlations (0.7-0.8). In fact, the value of 0.7 was indicated as the minimum acceptable value for achieving a relevant genetic progress in case of multiple-trait genomic predictions by Calus and Veerkamp (2011). Using this same criterion, NVF could reduce GHG emissions better than feed intake (AIF) given the stronger genetic correlations (0.74 vs ~0.65). Unsurprisingly, selection for larger animals will also lead to individuals that consume more feed with all correlations between growth traits and RIC-derived traits being close to 0.9. Similarly, selection for larger animals could lead to more GHG emissions (correlations close to 0.9). In general, selection indices could be built in order to reduce GHG emissions without compromising growth, condition, stature and feed intake. The upcoming research should involve the use of feed efficiency and relative GHG emission, as these components need to be adjusted by the growth and size of the animal.

Conclusion

Results suggest that selection indices could be built in order to reduce greenhouse gas emissions while still improving growth, condition, stature and feed intake. The upcoming research is involving the use of feed efficiency and relative GHG emission, as these components need to be adjusted by the growth and size of the animal as well as production records from cows that are sibs of the tested bull. Moreover, we will test these bull's semen in other experimental stations equipped with these precision instruments in order to register intake data and GHG data emission on some of the daughters of these bulls and re-estimate the genetic correlations between bulls and cows as proposed by other colleagues in the early nineties (Nieuwhof *et al*.,1992)

Table 1. Descriptive statistics (posterior means with posterior standard deviation) and heritability estimates for the traits analysed.

Trait[1]	Metric	N	Mean	SD	h^2
WEI	kg	885	309.3	77.5	0.45 (0.24)
BCS	score	849	3.0	0.3	0.51 (0.20)
HG	cm	715	157.3	14.2	0.44 (0.25)
HEI	cm	714	125.5	7.7	0.39 (0.23)
NVF	count	7150	26.0	11.6	0.31 (0.12)
AIF	kg	7150	0.3	0.1	0.17 (0.15)
ATF	s	7150	317.0	117.1	0.29 (0.18)
NVG	count	2817	3.9	1.7	0.36 (0.11)
CO_2	g/d	2817	6198.2	1103.9	0.48 (0.21)
CH_4	g/d	2817	223.6	51.8	0.40 (0.17)
AIR	L/s	2817	29.2	4.0	0.45 (0.09)
ATG	s	2817	329.3	87.5	0.24 (0.11)

[1] WEI: body weight; BCS: body condition score; HG: heart girth; HEI: height; NVF: number of visits at the feeder; AIF: average intake at the feeder; ATF: average time at the feeder; NVG: number of visits at the GreenFeed; CO_2 daily carbon dioxide emissions; CH_4: daily methane emissions; AIR: average airflow at the visit; ATG: average time at the GreenFeed.

Table 2. Estimates of genetic correlations[1] among the growth-related traits[2], the RIC-derived traits[2] and the GreenFeed-based traits[2].

	WEI	BCS	HG	HEI	NVF	AIF	NVG	CO_2	CH_4	AIR
WEI		**0.84**	**0.75**	0.64	**0.95**	**0.99**	**0.93**	**0.92**	**0.92**	**0.94**
BCS	**0.84**		**0.72**	0.55	**0.90**	**0.98**	**0.97**	**0.93**	**0.93**	**0.95**
HG	**0.75**	**0.72**		0.11	**0.90**	**0.98**	**0.94**	**0.90**	**0.90**	**0.94**
HEI	0.64	0.55	0.11		**0.90**	**0.97**	**0.95**	**0.92**	**0.92**	**0.95**
NVF	**0.95**	**0.90**	**0.90**	**0.90**		**0.75**	**0.73**	**0.63**	**0.67**	**0.69**
AIF	**0.99**	**0.98**	**0.98**	**0.97**	**0.75**		**0.67**	**0.55**	**0.58**	**0.61**
NVG	**0.93**	**0.96**	**0.94**	**0.95**	**0.73**	**0.67**		**0.70**	**0.77**	**0.92**
CO_2	**0.92**	**0.93**	**0.90**	**0.93**	**0.63**	**0.55**	**0.70**		**0.81**	**0.81**
CH_4	**0.92**	**0.93**	**0.90**	**0.92**	**0.67**	**0.58**	**0.77**	**0.81**		**0.83**
AIR	**0.94**	**0.95**	**0.94**	**0.95**	**0.69**	**0.61**	**0.92**	**0.81**	**0.83**	

[1] Estimates of the genetic correlations are the posterior mean for the parameter. Values in bold indicate that the value ‘0’ was not included within the 95% confidence intervals, therefore are to be considered significant for P<0.05.

[2] WEI: body weight; BCS: body condition score; HG: heart girth; HEI: height; NVF: number of visits at the feeder; AIF: average intake at the feeder; NVG: number of visits at the GreenFeed; CO_2 daily carbon dioxide emissions; CH_4: daily methane emissions; AIR: average airflow at the visit.

Acknowledgement

This study was supported by “Latteco2 project, sottomisura 10.2 of the PSRN Biodiversity 2020–2023” (MIPAAF. D.M. no. 465907 del 24/09/2021, project unique code 12C21004080005).

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al*. (2010). J. Dairy Sci. 93:2 743–752. https://doi.org/10.3168/jds.2009-2730.

Appuhamy J.A.D.R.N., France J. and Kebreab E. (2016). Glob. Chang. Biol. 22:9 3039–3056. https://doi.org/10.1111/gcb.13339.

Breider I.S., Wall E., and Garnsworthy P.C. (2019). J. Dairy Sci. 102:8 7277–7281. https://doi.org/10.3168/jds.2018-15909.

Callegaro S., Niero G., Penasa M., Finocchiaro R., Invernizzi G., Cassandro M. (2022). Italian Journal of Animal Science 21:1, 870-877, DOI:10.1080/1828051X.2022.2071178

Calus M.P.L. and Veerkamp R.F. (2011). Genet. Sel. Evol. 43:26. https://doi.org/10.1186/1297-9686-43-26.

Cassandro, M., Cecchinato, A., Battagin, M., Penasa, M., 2010. Genetic parameters of predicted methane production in Holstein Friesian cows. Page 181 in Proc. 9th World Congr. on Genetics Applied to Livestock Production, Leipzig, Germany.

Cassandro, M., 2013. Comparing local and cosmopolitan cattle breeds on added values for milk and cheese production and their predicted methane emissions. Animal Genetic Resources/Ressources gèneètiques animales/Recursos geneèticos animales, available on CJO2013. doi:10.1017/S2078 63361200077X.

Cassandro M, Mele M, Stefanon B. 2013. Genetic aspects of enteric methane emission in livestock ruminants. Ital J Anim Sci. 12:450–458.

de Haas Y., Pszczola M., Soyeurt H., Wall E., and Lassen J. (2017). J. Dairy Sci. 100:2 855–870. https://doi.org/10.3168/jds.2016-11246.

Gerber P.J., Hristov, A.N., Henderson B., Makkar H., Oh J., *et al*. (2013). Animal 7:2 220–234. https://doi.org/10.1017/S1751731113000876.

Haque M.N. (2018). J. Anim. Sci. Tech. 60:15. https://doi.org/10.1186/s40781-018-0175-7.

Hristov A.N., Oh J., Giallongo F., Frederick T., Weeks, H., *et al*. (2015). The use of an automated system (GreenFeed) to monitor enteric methane and carbon dioxide emissions from ruminant animals. JoVE J. Vis. Exp. 103, 52904. https://doi.org/10.3791/52904.

Lassen J. and Løvendahl P. (2016). J. Dairy Sci. 99:3 1959–1967. https://doi.org/10.3168/jds.2015-10012.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al*. (2002). Proc. of the 7th WCGALP. Montpellier, France.

Niero G, Cendron F, Penasa M, De Marchi M, Cozzi G, Cassandro M. 2020. Repeatability and reproducibility of measures of bovine methane emissions recorded using a laser detector. Animals. 10(4):606.

Pickering N.K., Chagunda M.G.G, Banos G., Mrode R., McEwan J.C. *et al.* (2015). J. Anim. Sci. 93:1 11–20. https://doi.org/10.2527/jas.2014-8302.

Nieuwhof G.J., van Arendonk J.A.M., Vos H., Korver S. (1992). Livestock Production Science 32:189-202

Estimates of genetic parameters for environmental efficiency traits for first lactation Holsteins

L.S.F. Lopes[1], K. Houlahan[1], F.S. Schenkel[1], D. Tulpan[1], F. Miglior[1,2], M. Sargolzaei[3,4], J.E. Koltes[5] and C.F. Baes[1,6]

[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences,
University of Guelph, N1G 2W1, Guelph, Ontario, Canada
Corresponding author: llopes@uoguelph.ca
[2]Lactanet, N1K 1E5, Guelph, Ontario, Canada
[3]Ontario Veterinary College, University of Guelph, N1G 2W1, Guelph, Ontario, Canada
[4]Select Sires Inc, 43064, Plain City, Ohio, United States of America
[5]Department of Animal Science, Iowa State University, 50011-3150,
Ames, Iowa, United States of America
[6]Institute of Genetics, Vetsuisse Faculty, University of Bern, 3012 Bern, Switzerland

Abstract

Rumination time was assessed as an indicator of efficiency and sustainability in dairy cattle. This study comprised 7782 records on 656 mid-first lactation Holstein cows. Animal models were used for rumination time and methane emission traits, while repeatability animal models were used for feed efficiency and production traits in bivariate analyses to estimate genetic parameters, including heritability, and genetic and phenotypic correlations between all traits. Rumination time had a moderate heritability (0.48±0.14) and genetic correlations of -0.45 (±0.25) with methane production, -0.88 (±0.24) with methane intensity, -0.08 (±0.19) with feed efficiency, and 0.48 (±0.18) with energy corrected milk. Although these findings should be validated in larger datasets, they suggest that rumination time has the potential to be used as an indicator trait for methane emissions and production levels.

Keywords: Rumination time, methane, dairy.

Introduction

Livestock is responsible for 6% of the global anthropogenic greenhouse gas (GHG) emissions (Gerber *et al*., 2013) with methane (CH_4) from the eructation of ruminants being a major contributor (Beauchemin *et al*., 2020). The production of CH_4 also corresponds to a 10% loss of dietary energy (de Haas *et al*., 2011). A decrease in CH_4 emissions would improve both efficiency and sustainability in the dairy sector. The recording of CH_4 and feed efficiency (FE) is costly and time-consuming, making the use of related traits that are inexpensive and easily measured necessary. One candidate indicator trait for CH_4 emissions and FE is rumination time (RT). Measured by automated sensors such as rumination collars, RT is already used at the commercial level in the monitoring of functional and production traits (Kaufman *et al*., 2018). For RT to be an indicator trait, it should be heritable and genetically correlated to traits of interest (Byskov *et al*., 2017). This study aimed to evaluate RT as an indicator of sustainability and efficiency in dairy cows by estimating genetic parameters, phenotypic and genetic correlations among RT, CH_4 emission, FE, and production traits in Canadian Holstein cows.

Material and methods

This study comprised 656 first lactation Holstein cows between 110 and 210 days in milk. Measured traits were rumination time (RT), methane production (CH_4), methane yield (MeY), methane intensity (MeI), dry matter intake (DMI), feed efficiency (FE), metabolic body weight (MBW), and energy corrected milk (ECM). Rumination time in minutes per day was obtained from Heatime® neck collars, (SCR HiTag, Allflex, Netanya, Israel) (Schirmann *et al*., 2009; Andreen *et al*., 2021). Milk samples were collected weekly and analysed by Lactanet Canada (Guelph, ON) for fat and protein content in kilograms. Energy corrected milk was calculated, where ECM = (0.25 x milk) + (12.2 x fat) + (7.7 x protein) (Sjaunja *et al*., 1990). Daily live body weight (BW) records were used for the calculation of MBW as $BW^{0.75}$. Records for DMI in kilograms were obtained as the product between total mixed ration (TMR) intake in kilograms and the calculated dry matter percentage in the diet, where TMR intake was calculated daily as the difference between offered and leftover feed. Feed efficiency was calculated by a recursive linear transformation of DMI based on the genetic (co)variances among DMI, ECM and MBW (Jamrozik *et al*., 2021). Methane emission was obtained using the GreenFeed® System (C-Lock Inc., Rapid City, South Dakota, USA) (Hristov *et al*., 2015; Huhtanen *et al*., 2015). Methane records were used to calculate MeY and MeI as grams of methane per kilograms of DMI and ECM, respectively. Each trait was assessed for outliers at three standard deviations from the average. All traits had to have a minimum of two records per week of lactation to be considered for analysis, where the week of lactation was defined using the milk test day. Finally, all traits were averaged for the week of lactation. Cows had repeated records of weekly averages for DMI, FE, ECM and MBW.

Statistical models

Animal models were used for RT, MeY and MeI, while repeatability animal models were used for FE, DMI, ECM, and MBW in bivariate analyses to estimate genetic parameters. All (co)variance components were estimated using the average information residual maximum likelihood algorithm in ASREML 4.0 (Gilmour *et al*., 2015). For each trait, heritability was the average result from all bivariate combinations. In general, the model used in this study was:

$$Y_{ijklm} = \mu + AC_i + WL_j + YS_k + a_l + pe_l + e_{ijklm},$$

where Y_{ijklm} represents the m^{th} phenotype (for RT, CH_4, MeY, MeI, FE, DMI, ECM, and MBW) of the l^{th} animal; μ is the overall mean of the trait; AC_i is the fixed effect of the i^{th} age at calving class in months (nine classes); WL_j is the fixed effect of the j^{th} week of lactation (fifteen levels); YS_k is the fixed effect of the k^{th} year and season of calving class (sixteen classes); a_l is the random additive genetic effect of the l^{th} cow; pe_l is the random permanent environmental effect (for FE, DMI, ECM and MBW) of the l^{th} cow, and e_{ijklm} is the random residual error term.

Results and discussion

Heritabilities

The heritability (h^2) estimates in Table 1 show that selection is possible for all analysed traits. The estimated RT h^2 (0.48±0.14) was larger than 0.33±0.16 and 0.34±0.05 previously reported (Byskov *et al*., 2017; Moretti *et al*., 2018). However, these studies used different trait definitions for RT, which could affect h^2 estimates. Estimated heritabilities of CH_4, MeI, and MeY are found to range from 0.05 to 0.38 (de Haas *et al*., 2011; Manzanilla-Pech *et al*., 2021) and the variation in the results can be attributed

Table 1. Genetic correlation (above diagonal), heritability (diagonal), and phenotypic correlation (below diagonal), for rumination time (RT), methane production (CH_4), methane yield (MeY), methane intensity (MeI), feed efficiency (FE), dry matter intake (DMI), energy corrected milk (ECM), and metabolic body weight (MBW).

	RT	CH_4	MeY	MeI	FE	DMI	ECM	MBW
RT (min/day)	**0.48 (0.14)**	-0.45(0.25)	NA	-0.88 (0.24)	-0.08 (0.19)	0.17 (0.14)	0.48 (0.12)	-0.24 (0.13)
CH_4 (g/day)	-0.10 (0.06)	**0.42 (0.12)**	0.85 (0.50)	0.48 (0.23)	0.13(0.18)	0.81(0.10)	0.76(0.14)	0.67(0.10)
MeY (g/kg)	NA	0.46 (0.04)	**0.12 (0.10)**	0.84 (0.62)	-0.91 (0.24)	-0.92 (0.12)	-0.37 (0.29)	0.04 (0.29)
MeI (g/kg)	-0.26 (0.06)	0.57 (0.04)	0.39 (0.05)	**0.36 (0.13)**	0.04 (0.22)	-0.17 (0.16)	-0.81 (0.08)	0.66 (0.13)
FE	0.05 (0.09)	0.02 (0.09)	-0.60 (0.06)	0.06 (0.09)	**0.13 (0.07)**	0.69 (0.14)	-0.06 (0.29)	-0.08 (0.30)
DMI (kg)	0.17 (0.07)	0.50 (0.05)	-0.70 (0.04)	-0.03 (0.07)	0.84 (0.01)	**0.24 (0.07)**	0.56 (0.17)	0.40 (0.21)
ECM (kg)	0.21 (0.07)	0.37 (0.06)	-0.06 (0.07)	-0.69 (0.04)	-0.07 (0.03)	0.34 (0.03)	**0.32 (0.07)**	-0.01 (0.21)
MBW ($kg^{0.75}$)	-0.17 (0.08)	0.45 (0.07)	0.01 (0.08)	0.43 (0.08)	-0.15 (0.04)	0.22 (0.04)	-0.03 (0.04)	**0.44 (0.11)**

to the environmental impact on the methane produced by cows, and the method of measurement (López-Paredes *et al*., 2020). The h^2 for FE (0.13±0.07) was within the reported range for FE (0.01 to 0.40) (de Haas *et al*., 2011; Vallimont *et al*., 2011). Thus, improvements in FE could be possible through selection, however, correlation with other economically important traits should be assessed.

Genetic correlations

Genetic correlations were estimated between all traits (Table 1). The bivariate analysis between MeY and RT did not converge, and this may have been caused by the model not being robust enough to estimate parameters for all traits in a small data set. Rumination time was uncorrelated to FE and had negative correlations with CH_4 (-0.45±0.25), MeI (-0.88±0.24) and MBW (-0.24±0.12) while showing positive correlations with the remaining traits. The genetic correlations between RT and ECM (0.48±0.12) could be a relationship to be exploited by selection programs. Greater rumination and production could be linked to a greater intake by the cows, as shown by the correlations between RT and DMI (0.17±0.14), and between ECM and DMI (0.55±0.16). Additionally, the increased intake could suggest a faster passage rate, and lower fermentation rate (Ramin and Huhtanen, 2013), possibly explaining the association between RT and $CH_{4,}$ and between RT and MeI.

Conclusion

The goals of this study were to estimate genetic parameters for automatically recorded RT and identify if RT is genetically correlated to efficiency, methane production, yield and intensity, and milk production. Our findings show that RT is heritable and is a candidate trait for the identification of low-emitting and high-producing animals with no direct impact on efficiency.

Acknowledgement

The authors acknowledge partners and collaborators of the efficient dairy genome project (https://genomedairy.ualberta.ca/) and the resilient dairy genome project (http://www.resilientdairy.ca).

References

Andreen, D.M., M.M. Haan, C.D. Dechow, and K.J. Harvatine. 2021. Determination of relationships between rumination and milk fat concentration and fatty acid profile using data from commercial rumination sensing systems. J Dairy Sci (104) :8901–8917. doi:10.3168/jds.2020-19860.

Beauchemin, K.A., E.M. Ungerfeld, R.J. Eckard, and M. Wang. 2020. Review: Fifty years of research on rumen methanogenesis: Lessons learned and future challenges for mitigation. Anim (14): S2–S16. doi:10.1017/S1751731119003100.

Byskov, M. v, A. Fogh, and P. Løvendahl. 2017. Genetic parameters of rumination time and feed efficiency traits in primiparous Holstein cows under research and commercial conditions. J Dairy Sci (100):9635–9642. doi:10.3168/jds.2016-12511.

Gerber, P.J., A.N. Hristov, B. Henderson, H. Makkar, J. Oh, C. Lee, R. Meinen, F. Montes, T. Ott, J. Firkins, A. Rotz, C. Dell, A.T. Adesogan, W.Z. Yang, J.M. Tricarico, E. Kebreab, G. Waghorn, J. Dijkstra, and S. Oosting. 2013. Technical options for the mitigation of direct methane and nitrous oxide emissions from livestock: a review. Anim (7) Suppl 2:220–234. doi:10.1017/S1751731113000876.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J.R. Welham, R. Thompson, 2015. ASReml user guide release 4.1 functional specification. Hemel Hempstead, UK: VSN International Ltd. Available at: https://www.vsni.co.uk

de Haas, Y., J.J. Windig, M.P.L. Calus, J. Dijkstra, M. de Haan, A. Bannink, and R.F. Veerkamp. 2011. Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. J Dairy Sci (94):6122–6134. doi:10.3168/jds.2011-4439.

Hristov, A.N., J. Oh, F. Giallongo, T. Frederick, H. Weeks, P.R. Zimmerman, M.T. Harper, R.A. Hristova, R.S. Zimmerman, and A.F. Branco. 2015. The Use of an Automated System (GreenFeed) to Monitor Enteric Methane and Carbon Dioxide Emissions from Ruminant Animals. J Vis Exp: JoVE (2015): 52904. doi:10.3791/52904.

Huhtanen, P., E.H. Cabezas-Garcia, S. Utsumi, and S. Zimmerman. 2015. Comparison of methods to determine methane emissions from dairy cows in farm conditions. J Dairy Sci (98): 3394–3409. doi:10.3168/JDS.2014-9118.

Jamrozik, J., G.J. Kistemaker, P. Sullivan, B.J. Van Doormaal, T.C.S. Chud, C.F. Baes, F.S. Schenkel, and F.M. Miglior. 2021. Genomic Evaluation for Feed Efficiency in Canadian Holsteins. Page in 2021 Virtual ICAR-Interbull. Interbull Centre, Leeuwarden, Netherlands.

Kaufman, E.I., V.H. Asselstine, S.J. LeBlanc, T.F. Duffield, and T.J. DeVries. 2018. Association of rumination time and health status with milk yield and composition in early-lactation dairy cows. J Dairy Sci (101): 462–471. doi:10.3168/jds.2017-12909.

López-Paredes, J., I. Goiri, R. Atxaerandio, A. García-Rodríguez, E. Ugarte, J.A. Jiménez-Montero, R. Alenda, and O. González-Recio. 2020. Mitigation of greenhouse gases in dairy cattle via genetic selection: 1. Genetic parameters of direct methane using noninvasive methods and proxies of methane. J Dairy Sci (103): 7199–7209. doi:10.3168/JDS.2019-17597.

Manzanilla-Pech, C.I.V., P. Loevendahl, D. Mansan Gordo, G.F. Difford, J.E. Pryce, F. Schenkel, S. Wegmann, F. Miglior, T.C. Chud, P.J. Moate, S.R.O. Williams, C.M. Richardson, P. Stothard, and J. Lassen. 2021. Breeding for reduced methane emission and feed-efficient Holstein cows: An international response. J Dairy Sci (104): 8983–9001. doi:10.3168/JDS.2020-19889.

Moretti, R., M.P.G. de Rezende, S. Biffani, and R. Bozzi. 2018. Heritability and genetic correlations between rumination time and production traits in Holstein dairy cows during different lactation phases. J Anim Breed Genet (135): 293–299. doi:https://doi.org/10.1111/jbg.12346.

Ramin, M., and P. Huhtanen. 2013. Development of equations for predicting methane emissions from ruminants. J Dairy Sci 96:2476–2493. doi:10.3168/JDS.2012-6095.

Schirmann, K., M.A.G. von Keyserlingk, D.M. Weary, D.M. Veira, and W. Heuwieser. 2009. Technical note/ : Validation of a system for monitoring rumination in dairy cows. J Dairy Sci (92): 6052–6055. doi:10.3168/jds.2009-2361.

Sjaunja, L.O., L. Bævre, L. Junkkarinen, J. Pedersen, and J. Setälä. 1990. A nordic proposal for an energy corrected milk (ECM) formula. Page in International Committee for Recording the Productivity of Milk Animals

Vallimont, J.E., C.D. Dechow, J.M. Daubert, M.W. Dekleva, J.W. Blum, C.M. Barlieb, W. Liu, G.A. Varga, A.J. Heinrichs, and C.R. Baumrucker. 2011. Short communication: Heritability of gross feed efficiency and associations with yield, intake, residual intake, body weight, and body condition score in 11 commercial Pennsylvania tie stalls. J Dairy Sci (94): 2108–2113. doi:10.3168/JDS.2010-3888.

Genetic correlation between methane emission and nitrogen use efficiency proxies in Walloon dairy cows

H. Atashi[1,2], Y. Chen[1], C. Grelet[3], A. Vanlierde[3], F. Dehareng[3], S. Vanderick[1], H. Soyeurt[1] and N. Gengler[1]

[1]Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium
Corresponding Author: nicolas.gengler@uliege.be
[2]Shiraz University, Shiraz, Iran
[3]Walloon Agricultural Research Center, 5030 Gembloux, Belgium

Abstract

Livestock production contributes to global human-induced greenhouse gases (GHG) emissions in the form of CH_4, N_2O and CO_2. Beyond climate change, direct CH_4 emissions are also linked with efficiency in dairy cows. The aim of this study was to estimate genetic correlations between methane emissions and nitrogen use efficiency (NUE) in Walloon dairy cows. Milk mid-infrared (MIR) spectra were used to predict CH_4 emissions (PME, g/d) using existing equations. The trait log-transformed CH_4 intensity (LMI) was defined as the log-transformed ratio of PME divided by the daily milk yield (MY, kg/d) recorded on the same test-day achieving a more normal distribution. The values of predicted NUE (PNUE) and N losses (PNL) as proxies of the NUE and N loss were obtained using the combined MIR spectra, parity, and milk yield-based prediction equations on test-day MIR records with days in milk (DIM) between 5 and 50 d. The used data were restricted to the first-parity cows. Random regression test day models were used to estimate genetic parameters with the Bayesian Gibbs sampling method using a single chain of 100,000 iterates with a burn-in period of 20,000 iterates. Mean (SD) daily h^2 estimated for PME and LMI were 0.14 (0.05) and 0.24 (0.05), respectively. Mean (SD) daily genetic correlation estimated between PME and LMI was 0.55 (0.03). At a level of reliability of more than 0.30 for all examined traits, breeding values of 420 bulls born after 1995 were used to estimate the approximate genetic correlations (AGC) between PME and LMI and PUNE and PNL. The AGC estimated between PME and PUNE was -0.33 (0.07) and that found between LMI and PUNE was -0.60 (0.07). The AGC estimated between PME and PNL was 0.43 (0.08) and that found between LMI and PNL was 0.32 (0.08). The results showed that CH_4 emission, as an indicator of energy lose, is positively correlated with predicted nitrogen lose and negatively correlated with N use efficiency. It can be concluded that genetic selection for decreasing CH_4 emission will also decrease N loss and increase N use efficiency in dairy cows.

Keywords: greenhouse gas emission, mid-infrared spectrum, genetic relationship

Introduction

Efficiency traits in dairy production have received increasing attention as affect not only farm profitability but also losses to the environment (Phuong *et al*., 2013). Therefore, the dairy industry is under constant pressure for further improvement of efficiency traits (Connor, 2015). If the nutrients consumed are not converted to milk, meat, body reserves, or new born calf, they are excreted into the environment, resulting in emissions such as urea (nitrogen (N) loss) and methane (energy loss) (Phuong *et al*., 2013). Proposed CH_4 emissions traits include CH_4 production, defined as the daily production (g/d); CH_4 yield, defined as the amount of CH_4 produced per unit of dry matter intake; CH_4 intensity, defined as CH_4 produced per unit of product (milk or meat), and residual CH_4. An increasing number of studies showed that N use efficiency (NUE), N loss (urinary urea, milk urea, and faecal urea), and CH_4 emissions are heritable traits and are correlated with milk yield and composition in dairy cows (Pickering *et al*., 2015; Chen *et al*., 2021a; Chen *et al*., 2021b; Richardson *et al*., 2021). However, both types of traits (NUE and CH_4 emissions) are challenging to record and to analyse on phenotypic and genetic level. Moreover, there is no report on the genetic correlation between proxies for N efficiency traits and methane emissions. Therefore, the aim of this study was to estimate approximate genetic correlations between CH_4 emission and the N use efficiency in Walloon dairy cows.

Material and methods

Data

Milk samples were collected on first parity cows from 2006 to 2021 by the Walloon Breeding Association (Ciney, Belgium). All milk samples were analysed using Milkoscan FT4000, FT6000, and FT+ (Foss-Electric A/S, Hillerød, Denmark) by the milk laboratory Comité du Lait (Battice, Belgium) to generate the MIR spectral data. Methane emissions (PME, g/d) were predicted from the recorded spectra using the equations developed by Vanlierde *et al*. (2021). To eliminate potential abnormal records, the PME values below the 0.1 percentile and above the 99.9 percentile were deleted (Kandel *et al*., 2017). Methane emission intensity (PMI, g/ kg of milk) was defined as the ratio of PME divided by the daily milk yield (kg/d) recorded on the same test-day. The PMI was then log-transformed to be normally distributed and called log-transformed CH_4 intensity (LMI). Records from days in milk (DIM) lower than 5 d and greater than 365 d were eliminated. Age at the first calving (AFC) was restricted to the range of 540 to 1200 d. The final dataset consisted of 1,529,282 test-day records on 229,465 first-parity cows distributed in 1,530 herds collected from 2006 to 2021. The EBV and reliability of predicted NUE (PNUE) and N losses (PNL) were obtained from (Chen *et al*., 2021b).

Variance component estimation and estimated breeding values

The (co)variance components for PME and LMI were estimated using the following random regression test-day model (RR-TDM).

$$y_{ijklm} = \mu + HTD_i + \sum_{b=0}^{4} AS_j ø_b(t) + \sum_{b=0}^{2} HY_k ø_b(t) + \sum_{b=0}^{2} pe_l ø_b(t) + \sum_{b=0}^{2} a_l ø_b(t) + e_{ijklm} \quad (1)$$

where $y_{ijyklmn}$ is the test-day record (PME or LMI) on DIM m of cow l, belonging to ith class of HTD, jth class of AS, and kth class of HY, HTD is the fixed effect of herd-test-day, AS is the fixed effect of age-season of calving defined as following: age at calving class (10 classes) × season of calving (four seasons),

$$\textstyle\sum_{b=0}^{4} AS_j ø_b(t)$$

is the fixed regression coefficients of the age-season at calving modelled using Legendre polynomials of order 4,

$$\textstyle\sum_{b=0}^{2} \mathrm{HY_k}\varnothing_b(t),\ \sum_{b=0}^{2} \mathrm{pe_l}\varnothing_b(t) \text{ and } \sum_{b=0}^{2} \mathrm{a_l}\varnothing_b(t)$$

are, respectively, the random regression coefficients of herd-year at calving (HY), permanent environment, and additive effects modelled using Legendre polynomials of order 2, and e_{ijklm} is the residual effect.

The (co)variance components were estimated by Bayesian inference using the GIBBS3F90 software (Aguilar *et al*., 2018). Gibbs sampling was used to obtain marginal posterior distributions for the various parameters using a single chain of 100,000 iterates. The first 20,000 iterates of the chain were regarded as a burn-in period to allow sampling from the proper marginal distributions. Genetic (co)variances on each test day were calculated using the equation described by Jamrozik and Schaeffer (1997). Daily heritability was defined as the ratio of genetic variance to the sum of genetic, permanent environmental, herd-year calving, and residual variances at a given DIM. The vector of estimated breeding values of each animal l (EBV_l), which included daily EBV from all DIM (5 to 365) as estimated by multiplying the vector of predicted additive genetic random regression coefficients by the matrix of Legendre orthogonal polynomial covariates; that is $\mathbf{EBV_l = T\hat{a}_l}$**, l** where is the vector of predicted additive genetic coefficients for animal l, and **T** is a matrix of orthogonal covariates associated with the Legendre orthogonal polynomial functions. In addition, the total EBV of animal l was obtained by averaging the EBVs from day 5 to 365 as following:

$$\mathrm{EBV_l} = \sum_{i=5}^{365} \mathrm{EBV_i}.$$

Approximated genetic correlation

The approximate genetic correlations (AGC) between PMI, LMI, and PNUE, PNL were estimated using the strategy presented by Blanchard *et al*. (1983). Bulls born after 1995 (included) with more than 0.30 in reliability for EBV of all traits were selected. The method of calculating standard errors of the AGC was the same as Chen *et al*. (2021b). The total EBV and reliability of PMI and LMI were used in this part.

Results and discussion

Means (SD) for PME, PMI and LMI were 326.8 (67.65) g/d, 14.97 (5.70) g CH_4/kg of milk yield, and 2.64 (0.35), respectively. The mean PME found (326.8 g/d) was in the range reported by previous studies (Pickering *et al*., 2015; Kandel *et al*., 2017). Heritability estimates for PME and LMI were relatively stable across lactation, with a mean h^2 of 0.14 and 0.24, and peak h^2 of 0.21 and 0.30, respectively. Moderate h^2 were estimated for PME and LMI in agreement with the literature (De Haas *et al*., 2011; Kandel *et al*., 2017; Sypniewski *et al*., 2021). A total of 420 bulls were used to estimate AGC between the proxies for N efficiency traits and methane emissions. Most of them belong to the US (137), the Netherlands (111), and Canada (52). The average reliabilities (SD) of EBV for PME, LMI, PNUE, and PNL were 0.84 (0.16), 0.84 (0.16), 0.48 (0.15), 0.49 (0.15), respectively. Estimated AGC (SE) between PME and PNUE was -0.33 (0.07) and between LMI and PNUE was -0.60 (0.07). The AGC (SE) estimated between PME and PNL was 0.43 (0.08) and between LMI and PNL was 0.32-(0.08). Methane emissions, as an indicator of energy lose, are positively correlated with PNL and negatively correlated with PNUE. These results seem to suggest that selection of NUE and CH_4 simultaneously is possible.

Conclusion

This study suggests that genetic selection for decreasing CH_4 emissions could also decrease N loss and increase NUE in dairy cows. The AGC estimates found in this study could be helpful to setup future breeding goals. Furthermore, we only start to apprehend the complexity of holistic breeding goals accounting not only for feed and production efficiencies, and GHG emissions but also resilience and productivity.

Acknowledgement

The China Scholarship Council (Beijing) is acknowledged for funding the PhD project of Yansen Chen. Hadi Atashi acknowledges the support of the “Service Public de Wallonie” (SPW – ARNE, Belgium) through the WALLeSmart Project (D31-1392). The SPW -ARNE, the National Fund for Scientific Research (Brussels, Belgium), the EMR EEIG and different European Projects (FP7: Robustmilk – Grant agreement n° 211708, GreenhouseMilk – Grant agreement n° 238562, GplusE - Grant agreement n° 613689, and H2020: The SmartCow - grant agreement n° 730924, INTERREG NWE: OptiMIR NWE190G and HappyMoo NWE730) supported the research leading to the used MIR prediction equations. The content of the publication reflects only the view of the authors; the European Community is not liable for any use that may be made of the information contained in this publication. The use of the computation resources of the University of Liège – Gembloux Agro-Bio Tech (Gembloux, Belgium) provided by the technical platform Calcul et Modélisation Informatique (CAMI) of the TERRA Teaching and Research Centre, partly supported by the National Fund for Scientific Research (F.R.S.-FNRS, Brussels, Belgium) under Grants No. T.0095.19 (PDR “DEEPSELECT”) is also acknowledged.

References

Aguilar, I., S. Tsuruta, Y. Masuda, D. Lourenco, A. Legarra and I. Misztal. 2018. BLUPF90 suite of programs for animal breeding with focus on genomics. Proceedings of the world congress on genetics applied to livestock production, p 11-16.

Blanchard, P., R. Everett and S. Searle. 1983. Estimation of genetic trends and correlations for Jersey cattle. J. Dairy Sci. 66: 1947-1954.

Chen, Y., H. Atashi, S. Vanderick, R. Mota, H. Soyeurt, H. Hammami and N. Gengler. 2021a. Genetic analysis of milk urea concentration and its genetic relationship with selected traits of interest in dairy cows. J. Dairy Sci. 104: 12741-12755.

Chen, Y., S. Vanderick, R. Mota, C. Grelet and N. Gengler. 2021b. Estimation of genetic parameters for predicted nitrogen use efficiency and losses in early lactation of Holstein cows. J. Dairy Sci. 104: 4413-4423.

Connor, E. 2015. Invited review: Improving feed efficiency in dairy production: Challenges and possibilities. Animal 9: 395-408.

De Haas, Y., J. Windig, M. Calus, J. Dijkstra, M. De Haan, A. Bannink and R. Veerkamp. 2011. Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. J. Dairy Sci. 94: 6122-6134.

Jamrozik, J and L. Schaeffer. 1997. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. J. Dairy Sci. 80: 762-770.

Kandel, P., M.-L. Vanrobays, A. Vanlierde, F. Dehareng, E. Froidmont, N. Gengler and H. Soyeurt. 2017. Genetic parameters of mid-infrared methane predictions and their relationships with milk production traits in Holstein cattle. J. Dairy Sci.100: 5578-5591.

Phuong, H., N. Friggens, I. De Boer and P. Schmidely. 2013. Factors affecting energy and nitrogen efficiency of dairy cows: A meta-analysis. J. Dairy Sci. 96: 7245-7259.

Pickering, N. K., M. G. Chagunda, G. Banos, R. Mrode, J. McEwan and E. Wall. 2015. Genetic parameters for predicted methane production and laser methane detector measurements. J. Anim Sci. 93: 11-20.

Richardson, C., T. Nguyen, M. Abdelsayed, P. Moate, S. Williams, T. Chud, F. Schenkel, M. Goddard, I. van den Berg and B. Cocks. 2021. Genetic parameters for methane emission traits in Australian dairy cows. J. Dairy Sci.104: 539-549.

Sypniewski, M., T. Strabel and M. Pszczola. 2021. Genetic Variability of Methane Production and Concentration Measured in the Breath of Polish Holstein-Friesian Cattle. Animals 11: 3175.

Vanlierde, A., F. Dehareng, N. Gengler, E. Froidmont, S. McParland, M. Kreuzer, M. Bell, P. Lund, C. Martin and B. Kuhla. 2021. Improving robustness and accuracy of predicted daily methane emissions of dairy cows using milk mid infrared spectra. J. Sci. Food Agric.101: 3394-3403.

Advanced analytics create insight and supports the Danish dairy farmers

J. Frandsen, T. Hejgaard Sørensen and L.A. Nielsen

SEGES Innovation, Denmark

Abstract

Danish dairy farmers have access to one of the most comprehensive management systems in the world. Data is the key to most management decisions on the farm and with advances analytics, data is transformed into valuable insights. Below are four examples of advanced analytic solutions available to the Danish dairy farmers.

- Predictive forecasting models is used to predict the cell count for the coming 4 months. The prediction is based on more than 7 years of data, taking detected seasonality into account.
- Advanced statistical models are implemented several places for example to predict lactation curves for individual and groups of cows. Advanced statistical models are also used to create a graphical analysis tool showing a longevity profile and a claw trimming profile for individual herds.
- Key influencers are derived from regression analysis in a reproduction management tool available to dairy farmers. The identified key influencers are the individual factors with the most impact on reproduction results.
- AI solutions such as Optical Character Recognition (OCR) and speech recognition is implemented in mobile apps to simplify data entries for the Danish farmers.

Introduction

The Danish Cattle Database system is unique in the sense that data from sources within the whole dairy industry is registered in one place and therefore provides a huge amount of information available for both farmers, advisors, and researchers, see the overview in figure 1. Examples of some of the more important data contributors to the Danish Cattle Database are:

- The farmers.
- Dairies – daily amount of delivered milk and analysis.
- Slaughterhouses – weight and quality data on culled animals.
- Breeding companies.
- Veterinarian management system.
- Milking equipment management systems.
- Feeding equipment management systems.
- Sensor equipment management systems.

The data registered on animal level are movement data, milk recording data including milk quality analysis, reproduction events from heat to pregnancy, registration of disease, treatment and medicine consumption, slaughter data like quality and weight, sensor data like activity and health alerts and in some cases also daily milk weight from milking equipment. On group and herd level the database receives delivered milk incl. quality analysis and feed consumption each day.

The big amount of data at the representation across all areas in the dairy production gives a good foundation for solid data analysis to guide the farmer and the service sector around the farmer.

Examples for data analysis in the Danish Management system, DMS

In the following there is a short description of how data are uses in analyses of the production to guide the farmer towards optimizing the production. The main engine in the data analysis is the Data WareHouse, DWH, where the raw data are crouched and processed to enriched key figures and used in different reports and systems. Most of the reports are build in Microsoft Power BI, which offers motivation dynamic presentation of the data, and a continuedly development in functionality

Forecast of the milk production

With certified data on milk production from the milk recording on more that 90% of the dairy cattle and validated movement data and solid reproduction data SEGES has analysed data and developed a system to forecast the production. By using data on the individual cow, it is possible to make the forecast on the individual cow. These forecasts are used in different analyses from lactation analyses to forecast of the production for budget purposes

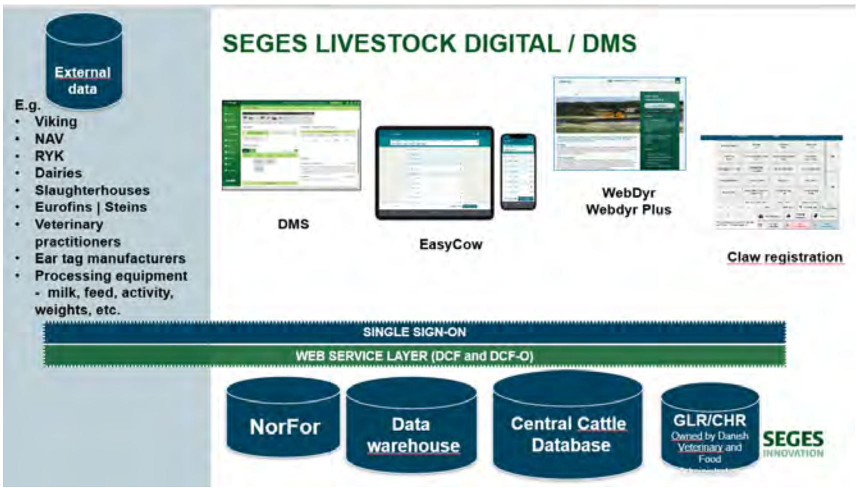


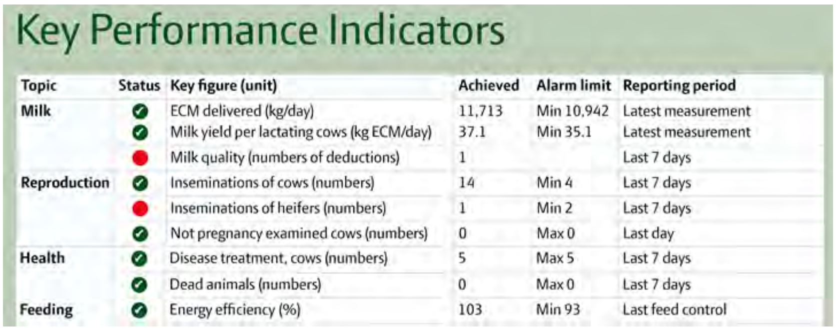
Figure 1. Overview of the Danish Cattle data, and management system.

Prescriptive data

The dairies in Denmark use SSC as a parameter in their payment to the farmer, which leads for a decrease in payment for e number of farmers. Bu using the forecast possibilities in Power BI on the data series in the database the farmer in DMS can see the consequences in payment id nothing is done. As a supplement to the data DMS present some advises about how to avoid the problem

Daily dashboard

By using data from the different sources DMS presents a daily Dashboard for the dairy farmer to guide in where to focus. It gives a daily KPI overview of the herd performance on milk production, reproduction, heath and feeding including a reference to goals for the performance.



Key Performance Indicators

Topic	Status	Key figure (unit)	Achieved	Alarm limit	Reporting period
Milk	✓	ECM delivered (kg/day)	11,713	Min 10,942	Latest measurement
	✓	Milk yield per lactating cows (kg ECM/day)	37.1	Min 35.1	Latest measurement
	●	Milk quality (numbers of deductions)	1		Last 7 days
Reproduction	✓	Inseminations of cows (numbers)	14	Min 4	Last 7 days
	●	Inseminations of heifers (numbers)	1	Min 2	Last 7 days
	✓	Not pregnancy examined cows (numbers)	0	Max 0	Last day
Health	✓	Disease treatment, cows (numbers)	5	Max 5	Last 7 days
	✓	Dead animals (numbers)	0	Max 0	Last 7 days
Feeding	✓	Energy efficiency (%)	103	Min 93	Last feed control

Benchmarking

A lot of farmers find a big motivation in benchmark against other similar herds performance. By using data and the daily data processing in the Data WareHouse DMS provides a Benchmarking system with more than 140 key figures with possibilities to benchmark against other comparable herds and own goals as seem in this example

Conclusion

All in all, the data in the Danish Cattle Database are used to analysis for the benefit of the farmer, it

- Gives the farmers a very qualified basis for making decisions on an operational and strategic level regarding their production.
- Gives the service entities around the famer the best possible basis for servicing and advising the farmers.
- Makes it attractive for the on farm equipment suppliers to join the data exchange for the benefit of the farmers.

Exchanges of French genetic data are driven by farmer consent managed in a dedicated blockchain

L. Journaux[1], L. Duveaux[2], H. Gilles[3] and S. Picardat[4]

[1]France Génétique Elevage (FGE), 149 rue de Bercy 75595 Paris cedex 12, France
Corresponding Author: laurent.journaux@france-genetique-elevage.fr
[2]Association SPIE, 9 avenue George V, 75008 Paris, France
[3]FAST, 149 rue de Bercy 75595 Paris cedex 12, France
[4]Agdatahub, 9 avenue George V, 75008 Paris, France

Abstract

FGE manages French collective system for the genetic improvement of ruminants and offers services to its members with shared genetic information systems. The information system for cattle genetic improvement manages 200 million of animal (+7.5 million per year) and more than 1 billion of information (+40 million per year), coming from 44 000 farms, recorded and used by more than 300 companies (135 with a very significant activity).

FGE launched in 2020 the deployment of a new professional national data base, connected with a dedicated data exchange platform and a breeder's consents verification platform "Agata Consent" to validate and secure the exchange based on a consortium blockchain solution. The system, in production for the first users, ensures the informed consent of breeders speaking about the use of their data by several professional organizations for research and development, genetic improvement or innovation, without any intervention from the user. It works as a trusted third party to exchange data even with multiple addressing, with addressing parameters supported by web services. It increases the confident of breeders and organizations in the exchanges and the uses of data all along the value chain, in adequation with the European Data Act regulation.

Introduction

Since the 60's, the French cattle breeders' organizations have been structuring their technical data recorded on farm for ruminant production (cattle, sheep and goat) at regional and national levels. Initially, the investment and infrastructures were funded and performed by state.

Gradually, the part of investment and cost supported by breeder's organization increased. Today, the cost and investments of the system are supported by farmer's funds for 90% and by state subsidies for 10%.

A first step occurred in 1998 to renew all the system in a distributed relational database. The system was well adapted to the land organization with, small to medium size, local farmer associations or cooperatives (Rehben *et al.*, 2005; Rognant *et al.*, 2005). The system, currently in production (Balvay 2017, Balvay *et al.*, 2014), has been managed by the national professional organization France Génétique Elevage (FGE) for 14 years with the development of new tools as consent management.

For Cattle, the information system covers animal identification, parentages, performances (i.e. milk production, weight, fertility and so on) and genotyping (DNA information) with all the technical information to use this data for statistical treatments (location, birth date, calving rank…). These data are specially used to estimate the breeding value of animal and to perform advices for the farmers; these data are exchanged at international level to establish international ranking of artificial insemination sires on breeding values. In addition, this information is available for farmers themselves and others clients like private companies.

The amount of data collected each year is very large as the total amount of information cumulated for 60 years.

The number of actors who contribute to the system is large: more than 300 data providers in 2019. 135 of them exchanged more than 10 000 records.

One specificity of this information system is to manage data on a very long term*:* the statistical studies performed to estimate breeding values needs a very large number of data and a knowledge of pedigree with a maximum of depth (minimum 4 generations of ancestors, i.e. data managed during a minimum of 30 years for Cattle!)

Table 1. Number of data managed in the cattle information system (in Millions)

	Identities	Births	Lactations	Milk records	Weighing	Morphology records	Inseminations
Records created in 2021	7,012	6,749	2,232	20,393	1,271	1,156	6,514
Total in the database	233,6	158,0	114,3	857,0	49,1	25,2	189,0

A new landscape for regulation

The data exchange in France are regulated by EU regulation on personal information – RGPD, CNIL 2022. In the next year this regulation will be updated by the Data Governance Act. While the Data Governance Act, presented in November 2020 and agreed by co-legislators in November 2021, creates the processes and structures to facilitate data sharing by companies, individuals and the public sector, it also clarifies who can create value from data and under which conditions. Especially, the Data Governance Act defines a new role of Data Intermediary Service Provider which must be independent of any storage or treatment of any kind. Furthermore, to ensure the liability and accessibility of data from production, beneficiary companies will have to ask for an explicit consent (prolonged GDPR for non personal data). This consent must be also verifiable on behalf of any data exchange. This new regulation will be in force by the end of 2023 (EU, 2022).

A new landscape of actor for new needs on data

Nowadays, the situation has evolved a lot compared to the beginning of the century. In some parts of the territory some small organizations remain structured around regional information systems managed by breeder cooperatives or associations dedicated to this technical activity. But, in other parts, big players (more than 100 M€ income) are individually part of the game. They have an internal information system covering all the aspects of the management of the company, technical data dedicated to genetic improvement is now only a part of information system structured in ERP (Enterprise Resource Planning).

The demand for exchange of data for multiple use is now very frequent and push the development of new solutions of exchange or sharing of data (iDDEn Reents and Pekeler, 2021; DataGene, Shaffer 2021; US Dairy Brain project Cabrera *et al.*, 2021.

Since 2017, FGE engaged a large project to adapt the existing system to this new deal. After 3 years of initial studies the project is entering into the final phase of development and production.

Description of the new system

The new system is a national database managed by FGE to capitalize all the data collected since the 60's and a set of web services to support data exchange with all the regional or private information systems.

The architecture of the new information system is organized to combine a national central data base with regional platform shared by several actors and private information systems where the technical information managed for genetic purposed is a part of integrated ERP of the company.

To take into account the mandatory respect of breeder consent in accordance with French and Eu regulation and to be compatible with policy of French Farmers Union, this new system is combined with a consent management system open to all the companies involved in cattle genetic improvement. This system is based on blockchain: this technology has been chosen for multiple reasons included its capacity of disintermediation and trust creation and the legal value of data encrypted and stored in the blockchain (as described in Leporcher, Goujon, Chouli, 2019)

Exchange between the different databases will be performed by web services

The development of a consent management system paired with a web services exchange system aims at ensuring the informed consent of breeders speaking about the use of their data by several professional organizations for research and development, genetic improvement or innovation. This system is implemented in the new organization to increase the confident of breeders and organization in the exchange and the uses of data.

- If an organization of genetic improvement environment obtained a consent of the farmer for a family of data, it can use the data in accordance with the use stated in the consent and it will have an easy access to the data using the tools developed.
- If a company outside the genetic improvement frame obtained a consent of the farmer and organization who recorded the data's, it can use the data in accordance with the use stated in the consent and it will have an easy access to the data using the tools developed.

The web service solution used is a very standard solution. The publication of web service messages will allow a large use by actors of cattle genetic improvement and other players.

To facilitate data exchange, and extension at national or European level, FGE seeks to maintain the highest compatibility with international data exchange standard. It's Includes, at the first place, the Animal Data Exchange Standard promoted by ICAR (ICAR ADE, 2022). This new French genetic data system already applies all the legal requirements concerning the Data Government Act.

Focus on Base Pro and exchange system

The Base Pro database is based on the PostgreSQL opensource technology.

Base Pro will allow data exchanges between all actors involved in ruminant breeding sector and extended to the livestock value-chain according to modalities compliant with IT state of the art: Web Services.

These tools with standardized interfaces increase the efficiency of access to data by economic players and ensure compliance with management rules for exchanges. In particular, integration into web services of the systematic consent verification before any access to data to check the existence of the data's right holder's consent is a key factor in the confidence and appropriation of the system by these actors..

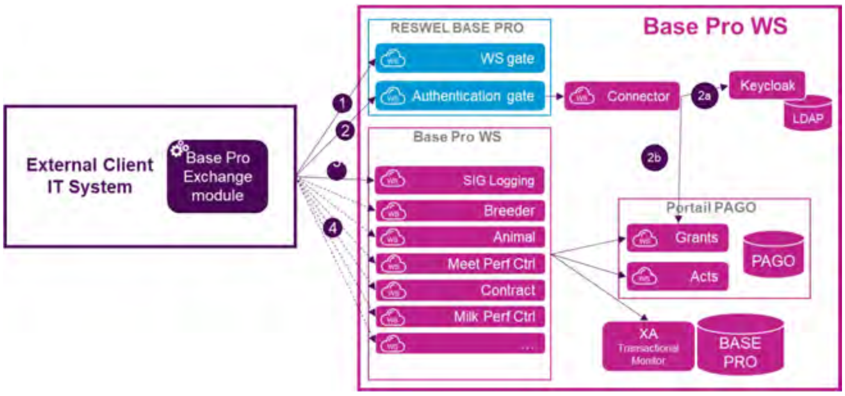


Figure 1. Base Pro Web Services functional architecture.

In the target architecture the client will access to the web services through the data exchange system (SED) as a trusted third party. The authentication is based on a Keycloak standard solution.

The technical application accounts are stored in a LDAP and managed through the administration portal PAGO which allows the management of the accounts and their access grants.

The technical framework is based on the standard technologies java/Rest/Wildfly/Swagger.

The Base Pro web services are deployed on a multi-node infrastructure which offers following features:

- Scalablity / Redundancy.
- Security.
- Multi-node load balancing and resilience.
- Continuous integration platform.
- Industrialized platform based on a standard components catalogue.
- Multi-version.

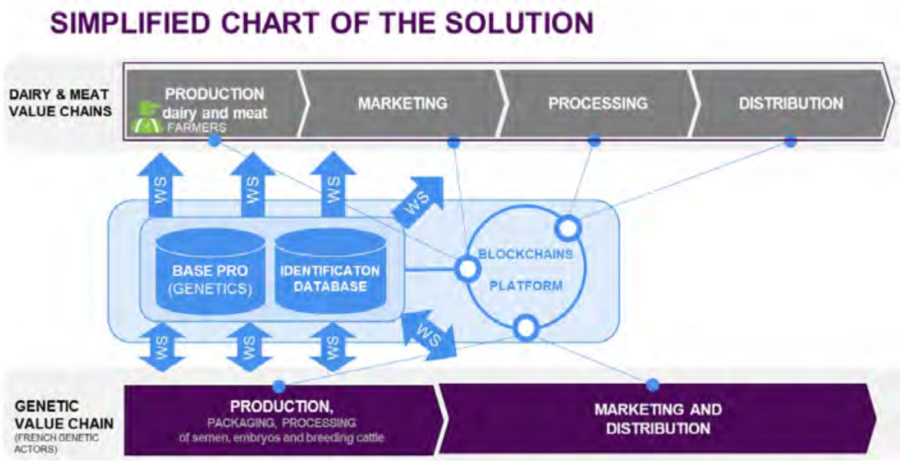


Figure 2. Simplified chart of solution.

- Documented Interface Contract (Swagger).
- Volumetric capacity:
  - 10 000 000 requests/month.
  - Peak : 20 000 000 requests/day.
  - Capacity Load Tests: 13 000 000 requests/hour.

Focus on consent management system

The consent management system called Agata Consent, developed by FAST company (FAST, 2022), was designed to achieve objectives of the Base Pro project: being able to ensure the collection of farmers' consent for data exchanges between various actors of the breeding value chain and giving them access to the consents they gave.

The system is combined with a data exchange management system (SED) developed by Agdatahub company (Agdatahub, 2022) to ensure security and trust.

Administrators of data domains / perimeters can access an ergonomic interface to administrate, create, modify the consents. They can also directly use the available APIs for the different functionalities. Thanks to a detailed rights gestion, farmers access the same interface to consult and accept / refuse consents.

Farmers have their own simple and pragmatic interface to consult and modify (according to contractual parameters) the delegations they gave to professional organizations and the related consents. This interface is exposed on Agri Maker (Agri maker, 2022), a digital services platform for agriculture, already gathering about 33 000 farmers and 25 services.

Agata Consent is a decentralized consent management solution based on open-source components that can be hosted on different cloud platforms using Kubernetes orchestrator. Webservices are developed based on Opensource Springboot framework. Full web management interfaces are developed with Javascript Angular framework, they are design responsive. Types encoding uses UTF-8 for exchange and storage, dates are encoded using the ISO 8601 norm.

Agata has a distributed architecture with a variable number of nodes, which is coherent with blockchain philosophy and ensure the security of the system by pseudonymization. Agata ensures consents privacy by encrypting all personal data in the blockchain.

Each node of Agata Consent Blockchain has the following technical architecture:

- HyperLedger Fabric Ledger*:* data storage in the blockchain.
- Smart contract*:* Java code used in the blockchain.
- Couch DB*:* database technology Fabric-CA*:* certification authority.
- Orderer*:* Hyperledger control.

The solution has a Tier III availability classification (with a maximum of 1,6 hours of unavailability per year)

If a specific organization obtained a consent of the farmer for a category of data (also called “data family”), it can use the data in accordance with the usage established in the consent. Moreover, the organization will have an easy access to the data using the tools developed. The system can also be used by a company outside the genetic improvement sector. Indeed, the Consent Management System is generically designed to be able to manage consents relating to data from all sectors of agri-food value chains and even other sectors of economy. It is organized into functional areas called “domains” within which data concerned and rules for managing consents are defined. This organization, which is both modular and generic, makes it possible to offer the service to a wide range of customers, while preserving specificities of each.

Agata Consent has been identified as a mean to obtain the label “Data Agri” for the companies who will use this solution.

For the system initialization, 273 071 existing consents have been loaded. It’s concern 74 534 herds and 138 companies.

Agata Consent is technically compatible with the French consent portal AgriConsent, (Agriconsent, 2022) the commercial version of Multipass project (Lauga *et al.*, 2019)

Conclusion

The first 2 domains on which the system combining SED and SGC will be deployed are ruminant identification and genetic improvement. Moreover, even if the system has been developed in this objective, it is applicable to all the agricultural value chain.

The initialization of the system ended in May 2022 and we are now int the phase of increase in load.

The promoters of this solution hope that it will be applied on a large fields of animal data exchange.

Reference

Agdatahub, 2022, https://agdatahub.eu/

Agriconsent, 2022,, https://agri-consent.eu/

Agrimaker, 2022, https://www.agri-maker.com/home.

Balvay B., 2017*,* A computerized consent management tool forbreeders: **why, how?** 1997. Proceedings of the ICAR Conference held in Edinburgh (UK) on 14-16 June 2017. M. Winters, M. Coffey, H. Newby, M. Burke and B. Wickham (Eds). ICAR Technical Series no. 22 (ISBN: 92-95014-18-9), p7-12

Balvay B., Rognant R. and Saunier D., 2014, new trends in data capture and exchange solutions between farms and external partners, French situation, ICAR Berlin; Friday May 23th 2014

Cabrera V.E., Fadul-Pacheco L., Wangen S., da Silva T., Zhang F., Fourdraine R.H. and Mattison J., 2021, Collecting, integrating, harmonizing and connecting data from dairy farms: the US Dairy Brain project experience, ICAR Technical Series #25, Circular farming and its impact on animal genetics, animal recording of data and cattle / herd management. Proceedings of the 44th ICAR Conference, virtually held from Leeuwarden, NL, 26-30 April 2021, Editors: G. de Jong, K. de Koning, H. van den Bijgaart. M. Burke and C. Mosconi, September 2021, p 215-219

CNIL, 2022, https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-onjn b

Data Agri, 2022, https://www.data-agri.fr/

EU 2022, European Commission, Data Act: Commission proposes measures for a fair and innovative data economy, 2022, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113; https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

FAST, 2022, https://www.fast-agritech.com/

ICAR ADE, 2022, https://www.icar.org/index.php/technical-bodies/working-groups/animal-data-exchange-wg/the-icar-ade-standards/

Lauga B., Balvay B., Topart L., Leclaire J., Clenet A., Brun F., Pinet F., Roussey C. and Sin M., 2019. MULTIPASS: managing the consents of access to farm data in a chain of trustto make new services emerge for farmers, Proceedings of the ICAR Conference held in Prague, CZ, 17-21 June 2019, New traits and adding new value to the recording and id services in the animal production Editors: J. Kucera, P. Bucek, D. Lipovsky, X. Bourrigan and M. Burk ePublisher: ICAR, 2019, p 177-184

Leporcher, Goujon, Chouli, 2019, Les Blockchains de la théorie à la pratique, de l'idée à l'implémentation, collection Epsilon, p18 and 115-120

Rehben E., Balvay B., Béréziat A., Bertrand C., Boulesteix I., Champy R. Dos C., Fusai B., Le Gal H, Marguin L., Poisnel E., 2005, Bilan de la rénovation du système d'information génétique français, Renc. Rech. Ruminants, 2005, 12, p 339-342

Reents R. and Pekeler T., 2021, iDDEN: An international perspective, ICAR Technical Series #25, Circular farming and its impact on animal genetics, animal recording of data and cattle / herd management. Proceedings of the 44th ICAR Conference, virtually held from Leeuwarden, NL, 26-30 April 2021, Editors: G. de Jong, K. de Koning, H. van den Bijgaart. M. Burke and C. Mosconi, September 2021, p 209-212

Rognant R., Breuil G., Gain G., Rehben E., Rouby A., Trocmé D., 2005, Les bases de données collectives d'élevage multi-partenaires*:* leurs apports aux systèmes d'information des éleveurs et des filières,, Renc. Rech. Ruminants, 2005, 12, p 319-322

Shaffer M., 2021, DataGene, Annual Update 2020/21, https://datagene.com.au/sites/default/files/inline-files/Datagene%20Annual%20Update%202021%20-%20FINAL.pdf, 28 pages

DCPO: The dairy cattle performance ontology, a tool for domain modelling and data analytics

V. Fuentes[12], T. Martin[1], P. Valtchev[1], A. B. Diallo[1,2], R. Lacroix[3] and M. Leduc[2]

[1]CRIA, Département d'informatique, UQÀM, Montréal (QC), Canada
Corresponding Author: diallo.abdoulaye@uqam.ca
[2]LACIM, Département d'informatique, UQÀM, Montréal (QC), Canada
[3]Lactanet, Sainte-Anne-de-Bellevue, Canada

Abstract

Dairy farming is being intensively computerized, whereby the goal is to use the recorded data to optimize production processes. This requires extensive analytics, which needs a good understanding of the data. It is also necessary that the datasets be federated to be able to get an integrated view. Although conventional database tools are helpful in that process, it is believed that linked data and ontologies can provide seamless integration of different sources while providing a semantic layer allowing deeper introspection of data. The objective was to build an ontology to provide such a semantic layer to dairy herd improvement (DHI) data.

A large dataset of milk production data was provided by Lactanet, Canadian Network for Dairy Excellence. This data is typically heterogeneous, i.e., covering partially or thoroughly health, nutrition, yield, and genetics. It also possesses a complex structure, with a large variety of data for a unique animal, dispersed in many records and multiple tables. A dedicated domain ontology, referred to as the Dairy Cattle Performance Ontology (DCPO), was built from a semantic analysis of the datasets. The initial core set of entities was determined using the definitions and minimal attribute sets for traits provided by ICAR guidelines and CDN documents. This core was gradually enriched with lower-level entities and aligned to more abstract concepts from the Basic Formal Ontology (BFO) to provide a foundational theory. The process was validated by domain experts. DCPO provides a rich and extensible data schema, a vocabulary based on international standards to support stakeholder collaboration. It federates external data sources and provides a semantic interface to query the obtained integrated linked data. Finally, DCPO underlies a knowledge base supporting analytics and decision making. Preliminary evaluation followed a query-based approach: SPARQL queries were designed reflecting typical questions experts might ask to assess the practical usability of DCPO.

Mining structural regularities, or patterns, in data may lead an expert to discover unknown phenomena or to confirm an already formulated hypothesis. The benefit of using DCPO as vocabulary for patterns is to enable seemingly unrelated yet isomorphic sub-graphs in the data with diverging vertex and edge labels, to become identical once their labels are generalized to DCPO classes and properties. Key benefit thereof was the patterns were described using the domain expert language to increase their interpretability. Next, we plan to use the ontology to support the deep learning-based inference of predictive models for milk production.

Keywords: Precision agriculture, dairy farming, domain ontologies, knowledge discovery from data, graph mining.

Introduction

Agriculture 4.0[1] refers to future trends helping the sector face the main challenges pertaining to the demands of the future. Precision farming, in particular, is about improving the overall farming process through in-depth analysis of its various aspects as reflected in their historical data generated by farming devices, produce/crop processing entities, regulatory bodies, etc. This requires all the stakeholders (e.g., producers, managers, analysts, consultants, etc.) to work together to leverage available data as a competitive advantage.

A typical approach is the design of machine learning or data mining-based analytical tools to, *inter alia*, predict outcomes in daily-life situations the stakeholders face or to detect major trends and/or exceptional events in the data. As living beings are involved, data are typically heterogeneous and complexly structured: They may cover such aspects as the well-being and health issues for farming animals, nutrition, yield, genetics, etc. Inner structure, e.g., time series, and inter-record relations, e.g., animal pedigree, would also appear in the data.

Constituting such complexly structured datasets requires a significant data-modelling effort. Moreover, as ever more aspects of the farming process get computerized, extensibility to further datasets is often a prime concern. This motivates a full-scale domain modelling in the form of a dedicated domain ontology (DO). DOs have a wide range of benefits beyond mere rich/extensible data schema. For instance, they provide a standardized vocabulary to support stakeholder collaboration while representing a centralized repository for domain expertise, thus enabling the design of decision-support systems for various domain tasks [1]. We present here the design of our *dairy cattle performance ontology* (DCPO), its current state and intended usage. The remainder of the paper is as follows: Section 2 presents our motivations while section 3 lists relevant prior work. Next, section 4 details our iterative modeling process and our tool set. Finally, section 5 concludes.

Motivation

Lactanet[2], is the dairy production centre of expertise covering the province of Quebec and the Atlantic regions of Canada. Lactanet's accumulated data about dairy production and milk control describes 6,670 herds and 1.5M cows. Key concepts reflected in the data include milk control samplings and the associated laboratory-based analyses that estimate the principal milk components: fat, protein, milk urea nitrogen, etc. Overall, the records provided by Lactanet amount to 3+ billion data end points. This huge dataset hides potentially meaningful concepts, e.g., unproductive cows admitting improvement vs those to quickly sell, and behavioral patterns for cows or farmers, that need to be uncovered. In order to allow richly structured heterogeneous datasets to be: (1) properly built and (2) analyzed to yield meaningful and intelligible patterns, we decided to design a DO. A number of our dairy analytical tools are symbolic-level, including graph mining methods whose cornerstone is a DO-powered generalized pattern miner. Additionally, a set of predictive models exploiting deep neural net architectures have been designed targeting a variety of yield metrics such as milk production and overall cost [2]. The way these can benefit from the ontology and the graph mining tools' output is currently under investigation.

---

[1] https://www.worldgovernmentsummit.org/api/publications/document?id=95df8ac4-e97c-6578-b2f8-ff0000a7ddb6

[2] https://lactanet.ca/en/home

Related work

Several ontological sources have been developed about dairy production. The *Animal Trait Ontology for Livestock* (ATOL, https://www.ebi.ac.uk/ols/ontologies/atol) models phenotypical animal traits. These are represented from an environment-aware and animal breeding-driven point of view. A *Common Dairy Ontology* (CDO) [3], has been designed towards assisting on farming decision making and semantic alignment (www.smartdairyfarming.nl).Yet CDO is primarily focused on sensor data and lacks a transverse view of the domain. *AgroRDF* [4] is a data exchange standard designed for agro-industrial purposes and built with semantic technologies. However, it lacks a unifying broader framework able to precisely describe the dairy domain. The *agriOpenLink* [5] system provides open interfaces and linked services to enable the development of new processes with a plug-and-play architecture. The *Dairy Farming Ontology* (DFO) is among the many created within the agriOpenLink project. Albeit strongly appealing for our own goals, it is not publicly available. The FAO (Food and Agriculture Organization of the United Nations) project develops agricultural standards such as *AgroVOC* vocabulary [6]. While it covers a wide range of subjects, it lacks middle-level concepts involved in dairy production, hence it is too generic for our needs.

In the recent past, DOs have been used to support a semantically rich data mining process. Indeed, they expose domain knowledge to machine processing while providing a rich vocabulary that is easily intelligible for domain experts [7]. Pattern mining [8], [9] aims at discovering recurrent data fragments in a dataset that might represent potentially useful trends and regularities (combinations of descriptors). Depending on data record topology and how much thereof is preserved in the patterns, various flavors of patterns have been studied, from itemsets (sets of products) to sequences to graphs. Independently, *generalized* patterns [10] have been introduced to deal with cases where abstracting from concrete data items (e.g. *Corona virus* instead of *SARS CoV 2* can bring insights absent in the ground level of data records. Generalized patterns are defined on top of an item taxonomy.

Graphs are among the most challenging pattern formats and adding a DO on top of their vertex and edge labels further compounds the issue. Partial solutions to the graph mining with a DO problem were investigated in [11]–[13]. Both [12] and [13] under-exploit the ontological structure by focusing only on parts of it (object properties and classes, respectively). In comparison, our DO is intended to support abstraction on edges as well, e.g. use *parent* property in patterns to match the *dam* property (female parent of a bovine) in data. In [11], abstraction from both vertices and edges was formalized, yet for graphs built around a vertex sequence which largely eases the mining task. In contrast, we deal with unrestricted graphs.

Finally, the problem of feeding the knowledge from a DO into a neural learning process was approached in [14] with class-embedding-based techniques. Prior studies have investigated mimicking the ontological structure by the neural network architecture [15]. Unlike these, we rely on discovered graphs patterns for data augmentation [16].

Building the dairy cattle performance ontology

We were provided with several non-ontological resources such as datasets of various provenance and coverage pertaining to dairy production, together with their data dictionaries. Additionally, we followed the International Committee for Animal Recording (ICAR, www.icar.org) guidelines and the publication of dairy cattle genetic evaluations in Canada provided by the Canadian Dairy Network (CDN, https://www.cdn.ca), consisting of publicly available data dictionaries. Starting from these resources, we applied an iterative modeling process inspired by the Ontology Summit 2013 Communiqué’s life cycle (://ontolog.cim3.net/OntologySummit/2013/communique.html. Below, we describe its main steps and outcomes.

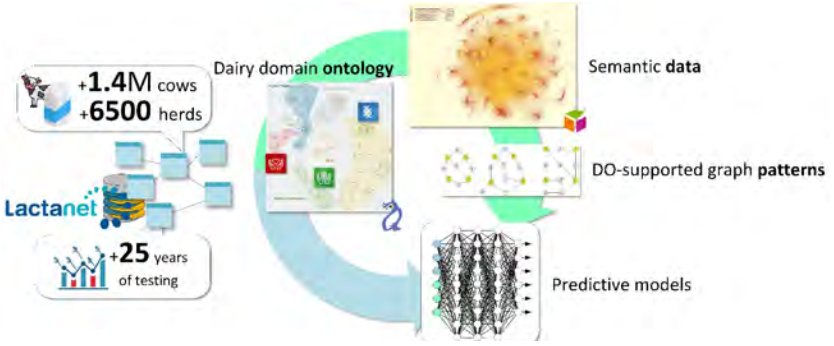


Figure 1. Global project architecture.

Requirements

Our initial focus was on purpose and scope of the ontology. Figure 1 depicts the global architecture of the information system the ontology is intended to support.

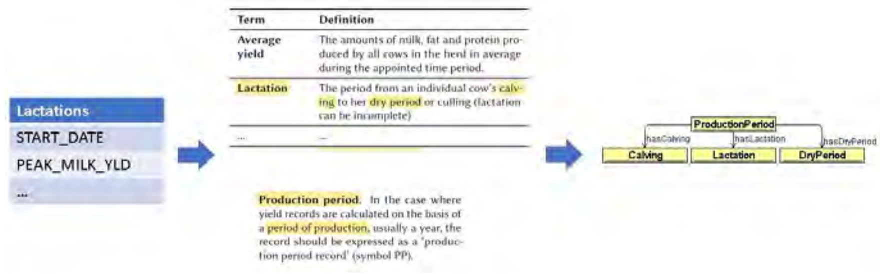
The DCPO, in the center, is intended to support three use cases: (1) a federated schema for external data sources on the left, (2) a semantic interface to cross-domain querying the linked data produced from the integration of external and internal sources, and (3) a knowledge base for graph mining algorithms on the right side. The produced generalized graph patterns are injected into further machine learning tools as additional ontology-based features to make their results more intelligible. As the results of this ontology supported system are intended to be used by anyone in the dairy community, it is important the compliance to international standards in the field, in particular to the ICAR guidelines and CDN's genetic vocabularies.

Scope

Precision dairy farming is about optimizing dairy cattle performance indicators that reflect complementary aspects of the dairy process. In the long term, DCPO will embody knowledge about six perspectives of the dairy process: breeding (pedigree), genetics, production and milk quality control, environment, health and nutrition. Currently, it encompasses only the first four, corresponding to the datasets and data dictionaries made available as departing point.

Ontological analysis and design

The ontological analysis for DCPO has been guided by domain experts, experienced data scientist, and the available structured description of the dairy data recording procedures within the ICAR documentation. This documentation was informative enough as to provide a core set of terms that was gradually enriched with lower-level concepts and properties and aligned to an upper ontology to provide a foundational theory.



Term	Definition
Average yield	The amounts of milk, fat and protein produced by all cows in the herd in average during the appointed time period.
Lactation	The period from an individual cow's calving to her dry period or culling (lactation can be incomplete)
...	—

Production period. In the case where yield records are calculated on the basis of a period of production, usually a year, the record should be expressed as a 'production period record' (symbol PP).

Figure 2. Bottom-up analysis: From dictionary names, via ICAR terms, to DCPO entities.

To take advantage of the available resources mentioned before, a bottom-up approach was performed. To identify the key entities of the ontology, we extracted the names of our datasets and their columns from available data dictionaries and matched them to the terms defined in the ICAR documents, so that we could link them to the standardized dairy domain terminology, Figure 2 illustrates this process. On the left, the candidate term *Lactation* is retrieved from the dataset name, and it's matched to the terms defined in the ICAR document, where an occurrence is found. In the matched definition, the related candidate terms/phrases *Calving* and *DryPeriod* are retrieved. Additional examination of the document identifies another candidate term *ProductionPeriod* and its relationship to the other terms are inferred (e.g. *hasLactation*).

In defining hierarchies (i.e., classes and properties) we usually provide an abstract level to factor out the common characteristics of the elements in a particular module, and one or more specialization levels below which inherit and refine these characteristics. This facilitates the management of the overall ontology architecture and inter-module connections, its extension, readability, and better grouping of similar entities. In some cases, the generalization process leads to the finding of *ontology patterns* (OP) that can then be reused across the ontology to provide modularity. Moreover, this decomposition of the ontology in abstraction layers allow the discovery of *generalized graph patterns* (GGP) from data, as we will see in the next section.

One such OP is the *ascertainment pattern* shown in Figure 3 (left). In detail, a target, *Thing*, undergoes an assessment procedure of some kind or *Ascertainment*, about some *DeterminableQuality* of the target and it is quantified by some measure or *DeterminateQuality*. This OP abstracts different ways of acquiring certain knowledge concerning the target entity. Under the umbrella of this OP, one finds such dairy farming activities as milk composition tests, genetic evaluations, and cow conformation scoring, to name a few. For instance, genetic evaluation is depicted on the right of Figure 3.

In searching for abstractions and OPs, we combine the bottom-up strategy of generalizing from concrete entities with the top-down strategy of making them specializations of a foundational ontology, the *Basic Formal Ontology* (BFO) from the OBO Foundry in our case. This greatly simplifies the integration of the two ontologies as the specialization approach gradually refactors the DCPO using BFO as a design guide, trying to align our entities to entities in the upper ontology. This has the effect of forcing our design to comply to the upper ontology, and thus absorb its principles. As an example, a genetic trait is any measurable characteristic of a cow that is heritable with some probability. Using the bottom-up strategy, we found a hierarchy of trait classes associated with concrete measures. From the top-down perspective

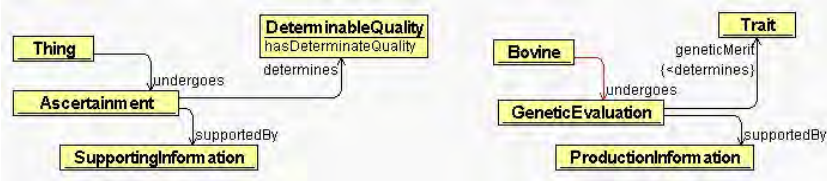


Figure 3. The OP Ascertainment (left) and its instantiation as GeneticEvaluation (right).

we understand that traits are *BFO*:*Quality* specializations. So, we created the classes *DeterminableQuality* and *DeterminateQuality* for general use in our patterns, which are both specializations of *BFO*:*Quality* and generalizations of our concrete classes for traits and measures correspondingly (see Figure 3).

This strategy enabled the rapid design of a coarse first model made of candidate classes and properties. We chose OWL since it is a standard Semantic Web technology built on top of RDF, a data format designed for interoperability, and provides valuable inference capabilities allowing ontology consistency to be checked with a reasoner, thus greatly reducing the production effort for the formal ontology artifact. Domain experts challenged this first model/design. This strategy enabled an analysis/design process following an iterative feedback loop (refining, updating, and adding new entities) with domain experts regularly challenging the latest changes.

Ontology description

The ontology has been modularized according to the different aspects of the dairy process covered at this stage: core, production and quality control, testing, breeding, and genetic evaluation. In the following paragraphs we summarized the ontology description as depicted in Figure **4**. Notice the use of italics to highlight ontology entities where classes begin with an uppercase letter and properties with a lowercase one.

At the core of the ontology, the central entity *Bovine*, factors out common characteristics of main actors: *Cow* and *Bull*, regardless of their particular role in the process or their life stage, allowing these concrete specializations to refine a common base by inheritance. *Bovine* is derived from *Animal*, used to enable extensions of the DCPO to other dairy species. A *parent* property and its specializations *femaleParent* and *maleParent* are defined on *Animal* to allow the construction of a parentage graph tracking the pedigree of each animal, with further specializations *dam* and *sire* for the cows and bulls involved in breedings, respectively.

The productive periods of a cow have three main stages: *Calving*, representing the birth of a new calf; *Lactation*, the milk production periods the cow has went through and *DryPeriod*, the time the cow is not producing milk. During *Lactation*, the *Milking* of cows undergoes *QualityControl* whose instances represent the different milk quality checkpoints performed during lactation. Quality control performs *MilkSampling* to produce a *MilkSample* that *isAnalyzedBy* a *CompositionTest*. During *MilkSampling* a *QuantificationTest* is performed to measure the milk yield . A *Breeding* between a *breedingDam* and a *breedingSire engenders* a new *Calving* producing a newborn *Bovine*. Each *Bovine*, undergoes a *GeneticEvaluation* that determines the *geneticMerit*

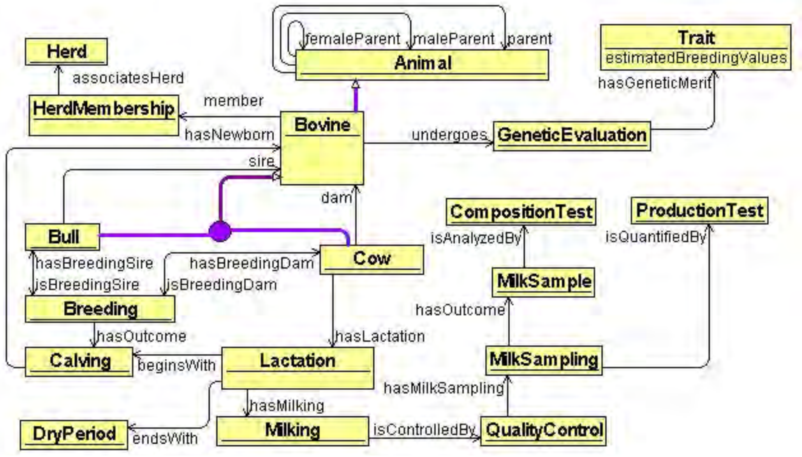


Figure 4. Dairy ontology and its modular desig.

of the animal on several *Traits*, to assess its value (a full description is available through CDN). Finally, a *Herd* entity is associated with the concept of *HerdMembership*, representing the fact that a cow belongs to a herd.

Ontology usage and evaluation

Query-based evaluation

As a preliminary evaluation of the DCPO, we adopted a query-based approach. The motivation behind was two-fold: (1) assess the practical usability of the populated ontology and (2) ensure the correctness of applied data transformations. Led by domain experts, we implemented SPARQL queries that reflect the typical questions experts might ask, e.g., to estimate the impact of cow management w.r.t to genetic potential. An example query is to compute average values on day 305 estimates for milk, protein, and fat for cows, herds, and regions for both production metrics (i.e., milk, fat, and protein) and estimated genetic potential (i.e., estimated breeding values). By substracting — relative to average — values for production and genetics, rough estimates of the quality of management practices for cows and herds are computed.

Generalized graph pattern mining with a DO

Structural regularities, or patterns, in the data can provide useful insights as to the general trends it reflects: They may lead an expert to discover unknown phenomena or, more realistically, to confirm an already formulated hypothesis. Therefore, such regularities, are worth mining and presenting to experts for an in-depth examination.

The immediate benefit of using a DO as vocabulary for pattern graphs is to enable the shared structure in data graphs to be explicitly described at the conceptual level, even though it may manifest in diverging ways at the data level. In other words, isomorphic graphs on the data level with diverging vertex and edge labels, which are thus, seemingly, unrelated, can become identical once their respective labels are generalized to the respective classes and generic properties from the DO.

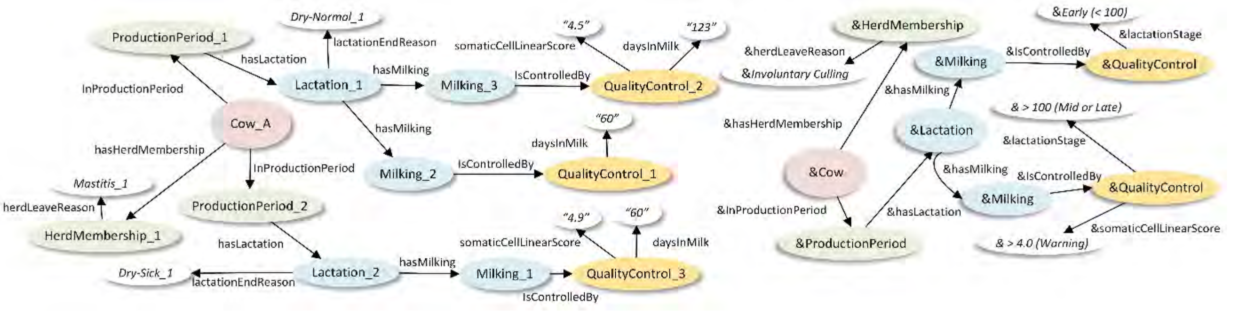


Figure 5. Example pattern (right) from a data graph (left), both supported by the dairy ontology.

Here, our DCPO and its instances act as a dual graph model where the former is used as a blueprint while the latter acts as the actual data to explore and analyze. Another way to picture it is to consider it as meta-data to formulate relevant hypothesis whereby graph data is used to (in-)validate such hypothesis.

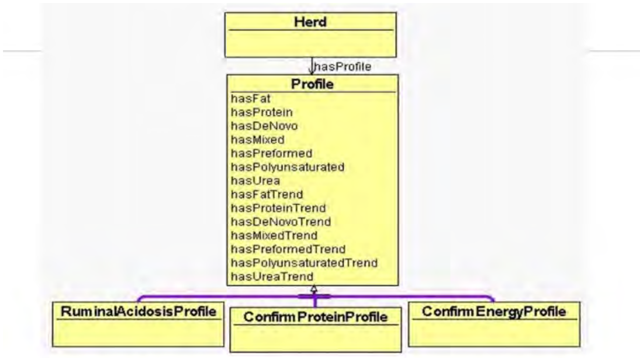
As an illustrative example, Figure 5 represents a data graph and a matching pattern that refers to DCPO. The pattern —found in an *ad-hoc* manner— was deemed potentially useful by our experts. It reflects the fact that a number of cows culled for reasons that were not under farmer's control (involuntary culling) had, prior to that event, at least one lactation with two quality controls, one of which indicates worrisome values of somatic cells. Such a co-occurrence is perfectly plausible as increased somatic cell counts are major signals for *mastitis* (inflammation of the udder tissue). Consequently, larger patterns contextualizing recurrent health issues could very well reveal the actual trigger for the involuntary culling. Therefore such patterns deserve to be investigated so that the underlying phenomena could be better understood and, if necessary, more closely monitored.

Experts' rules implementation

In the ontology, we describe what things exist in the domain, their attributes, and relationships. In general, this knowledge is external to animal experts, it is mostly established knowledge about the dairy process and its validity is not expected to change with the addition of new knowledge. But exists also internal knowledge developed by local experts. This internal knowledge usually expresses conditions on the state of individuals under which certain phenomena occurs. We use rules as knowledge representation in this case.

An example of this situation is the diagnosis of positive or negative effects of nutrition, management, and environmental factors in herds according to their bulk tank milk component profiles. The specification of the intervening entities (herd, milk profiles), their attributes (the values of the different measures for each profile component) and the relationships between them are expressed in the ontology (see Figure 6).

Animal experts have determined associations between extreme values of profile measures and different anomaly situations in herds. To express this associations, we define rules that use the vocabulary in the ontology. In this way, the ontology, rules, and a reasoner constitute a decision support system (DSS). Figure 6 shows an example of such rules: when a profile record is introduced in the ontology as an instance of a profile, the reasoner tries to match rule antecedents with instance data, if a match holds, the rule triggers and its consequent is executed, in this case it reclassifies the profile as a profile specialization.



```
hasFat=X and hasProtein=Y and... -> RuminalAcidosisProfile
hasUreaTrend=W and hasDeNovo=Z and... -> ConfirmEnergyProfile
...
```

Figure 6. Part of the ontology representing bulk tank milk profiles for herds.

Conclusion

We reflect here on our efforts on the design and implementation of DCPO, unifying several key aspects of dairy production. A major challenge we faced was the trade-off between plausible domain modeling and support for expressive knowledge discovery tools. At the current stage, it proved possible to reach both goals within a unique ontlogy. Next, we shall look at how to exploit ontology design patterns [18] and given we conform to the BFO, we envision an integration to the OBO (www.obofoundry.org) ontologies, with alignments to the relevant ontologies of the library. In longer run, we shall look at enhancing the data-centered ontology with knowledge discovered from the data by mining tools.

References

[1] **C.-J. Su and S.-F. Huang**, “Ontology-supported knowledge management with case-based reason-ing for intelligent health projection,” in BIBE, 2018, pp. 1–4.

[2] **C. G. Frasco and others**, “Towards an effective decision-making system based on cow profitability using deep learning,” in ICAART 2020, 2020, vol. 2, pp. 949–958.

[3] **J. P. C. Verhoosel and J. C. Spek**, “Semantics for big data applications in the smart dairy farming domain,” in Precision Dairy Farming Conference, Leeuwarden (NL), 2016.

[4] **D. Martini and others**, “agroRDF as a Semantic Overlay to agroXML: a General Model for Enhancing Interoperability in Agrifood Data Standards,” in CIGR Conference on Sustainable Agriculture through ICT Innovation, 2013.

[5] **S. D. K. Tomic and others**, “agriOpenLink: Towards Adaptive Agricultural Processes Enabled by Open Interfaces, Linked Data and Services,” in MTSR, 2013.

[6] **C. Caracciolo and others**, “The AGROVOC linked dataset,” Semant. Web, vol. 4, no. 3, pp. 341–348, 2013.

[7] **F. Kramer and T. Beissbarth**, “Working with ontologies,” in Bioinformatics, 2017, pp. 123–135.

[8] **C. C. Aggarwal and H. Wang**, Eds., Managing and Mining Graph Data, vol. 40. Springer, 2010.

[9] **C. C. Aggarwal and J. Han**, Frequent Pattern Mining, 2014th ed. Springer, 2014.

[10] R**. Srikant and R. Agrawal**, “Mining generalized association rules,” Futur. Gener. Comput. Syst., vol. 13, no. 2—3, pp. 161–180, 1997.

[11] **M. Adda and others**, “Toward {Recommendation} {Based} on {Ontology}-{Powered} {Web}-{Usage} {Mining},” Internet Comput. IEEE, vol. 11, no. 4, pp. 45–52, Aug. 2007.

[12] **T. Jiang and others**, “Mining generalized associations of semantic relations from textual web content,” IEEE TKDE, vol. 19, no. 2, pp. 164–179, 2007.

[13] A**. Cakmak and G. Ozsoyoglu**, “Taxonomy-superimposed graph mining,” in 11th EDBT, 2008, pp. 217–228.

[14] **U. Kursuncu and others**, “Knowledge infused learning (K-IL): Towards deep incorporation of knowledge in deep learning,” in AAAI-MAKE Symposium, 2020.

[15] **H. Wang and others**, “Ontology-based deep restricted boltzmann machine,” in DExA, 2016, pp. 431–445.

[16] **T. Martin and others**, “Bridging the gap between an ontology and deep neural models by pattern mining,” in CSSA@CIKM, CEUR Workshop Proceedings, 2020, vol. 2708.

[17] **M. Rouane-Hacene and others**, “Relational concept analysis: mining concept lattices from multi-relational data,” Ann. Math. Artif. Intell., pp. 1–28, 2013.

[18] **A. Gangemi and V. Presutti**, “Ontology design patterns,” in Handbook on ontologies, Springer, 2009, pp. 221–243.

Disease risk prediction based on an integrative data-methodological approach in dairy cattle

J. Lasser[1], C. Matzhold[2], C. Egger-Danner[1], F. Steininger[1], B. Fuerst-Waltl[4], T. Wittek[5] and P. Klimek[2]

[1]Graz University of Technology / Institute of Technical Informatics, Inffeldgasse 16/I,
8010 Graz / Complexity Science Hub Vienna, Josefstaedter Strasse 39,
1080 Vienna, Austria
[2]Medical University Vienna / Complexity Science Hub Vienna,
Josefstaedter Strasse 39, 1080 Vienna, Austria
[3]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89, 1200 Vienna, Austria
[4]Department of Sustainable Agricultural Systems, Division of Livestock Sciences,
University of Natural Resources and Life Sciences (BOKU),
Gregor-Mendel-Str. 33, 1180 Vienna, Austria
[5]Clinic of Ruminants, University of Veterinary Medicine,
Veterinaerplatz 1, 1210 Vienna, Austria
Corresponding author: egger-danner@zuchtdata.at

Abstract

Cattle breeding has been data driven since decades ago, with central data storage and data processing. Breeding goals are including more and more trait complexes and more and more diverse data sources have been recorded over time. Pedigree and genomic information in combination with a variety of phenotypes for dairy, beef, functional and conformation traits are available. The increasing awareness and need for transparency in regard to health and welfare has led to further documentation and data sources as well as the further exploration of existing data sources, e.g. DHI with mid-infra-red spectra. The need to increase efficiency has raised interest in feeding and management information and technological advances are revolutionizing the dairy sector with a large amount of novel data sources. One of the aims in the D4Dairy project is to explore the joint use of diverse data in combination with machine learning methods disease risk prediction and prevention.

Based on information from 165 farms with data from different life domains such as milk recording, genetic, housing, nutrition, management, climate and health , algorithms to predict the disease risk for lameness, acute and chronic mastitis, anestrus, ovarian cysts, metritis, ketosis (hyperketonemia), and periparturient hypocalcemia (milk fever) have been derived. The results are encouraging, as, for example lameness can be predicted with high sensitivity and specificity (F1=0.74). Using three machine learning approaches of varying complexity (from logistic regression to gradient boosted trees) it was shown that to some extent the complexity of the algorithm can compensate for less diverse data. Presently the studies are ongoing with focus on elaboration of a data-driven decision support tool for early warning to reduce the disease risk.

Keywords: Big data, disease risk prediction, data integration.

Introduction

Due to technological progress a huge amount of different data is generated in dairy operations. Data and central data processing have been a main pillar for breeding for decades. Breeding based on data and central data processing has started more than fifty years ago. Data are generated from many different data sources on and off the farm. As breeding goals have expanded in recent decades, interest in linking various existing data sources has increased. These include traditional data from performance recording or linear scoring of animals, as well as veterinary diagnoses, data from claw trimming or data from laboratories. Since about 15 years ago, genomic information based on single-nucleotide-polymorphisms (SNP) is used in breeding and more and more genomic information is available. The advances from research on mid-infrared-spectra (MIR) show potential to use this information for early warning of diseases or optimization of feeding. The advances in sensor technology is another big step forward. More insight into animal physiology and behavior is possible, as well as more detailed information from the milking process. Feeding information, sensors measuring the housing climate in the stable as well as other information describing environmental conditions offer new possibilities for developing decision support tools to predict an animal's risk of disease.

Digitization offers many opportunities to develop new tools as well as to improve the existing approaches in high-efficiency dairy farming. Nevertheless, the challenge very often is the lack of standardization and linkage of data.

In the project D4Dairy, a highly integrated dataset with data from the various data sources mentioned above was created and a holistic approach for prediction of diseases was developed.

The research questions are based on the assumption that digitization has the potential to significantly improve early detection and prevention of animal diseases. With our research, we wanted to answer to main research questions

- Do complex integrated datasets enable new approaches to the detection of animal diseases or improve existing ones.
- Do advanced methodological approaches improve prediction performance?

Material and methods

The prerequisite to study these research questions is an integrated dataset fulfilling different requirements with regards to data quality and quantity as well as the methodological approaches for the analyses of large data sets.

The dataset used for this study was an existing data set from the project "Efficient Cow" (Egger-Danner *et al*., 2017) consisting of data from 166 farms, with 142 different variables, 6,519 cows and 45,944 observations. In addition, data from the national disease registry and national weather service were combined with the existing data set. The data set included data on diagnoses, housing, breed, age, management, conformation, feed, breeding values, lactation stage, environment and milk performance. The study concentrated on the eight most frequent diseases lameness, acute and chronic mastitis, anestrus, ovarian cysts, periparturient hypocalcemia, ketosis, metritis (Lasser *et al*. 2021).

In order to get more insight into the development of disease, as it is often an interaction of multiple factors, farm-risk-profiles were derived (Matzhold *et al*. 2021). A UMAP (Unifold Manifold Approximation and Projection) algorithm for dimension reduction and to identify pertinent factors was used. The HDBSCAN algorithm was used to identify clusters of farms with similar attributes. With multivariate regression models the impact of individual risk factors was analysed. Each farm cluster identified a farm as a combination of environmental factors and management practices.

To predict the disease risk, three methods with increasing complexity were studied. The baseline was a logistic regression approach. In addition, two more complex machine learning approaches (random forests and gradient boosted trees – XGBoost) were employed to assess the improvements in disease prediction performance with increasing algorithm complexity. The methods were applied on two different datasets: A full dataset including information from the “Efficient Cow” project and weather information, and a reduced dataset that was comprised of data that is routinely collected by performance recording organisations.

Results and discussion

Figure 1 shows the farm-risk-profiles for the eight different diseases (Matzhold *et al*. 2021). Cluster 1 is characterized by the lowest prevalence of anestrus, ketosis, chronic mastitis, metritis and ovarian cysts. These are mainly farms in higher altitudes and with access to pasture. Cluster 4 showed the highest prevalence in anestrus, acute mastitis, lameness and ovarian cysts. These farms are characterized e.g. by less individualized feeding.

In Table 1 the F1-Score, Precision and Recall for predicting the eight diseases are shown for the logistic regression applied on the full dataset (Lasser *et al*., 2021). Precision is defined as probability that a predicted disease will actually be diagnosed, recall is the probability that an actual disease was correctly predicted. The F1-Score is the harmonic mean of precision and recall. Prediction works well for anestrus and lameness, with an F1-Score of 0.74. Prediction performance is worst for acute and chronic mastitis, with F1-Scores of 0.48 and 0.51, respectively. It has to be considered that the dataset collected did not include specific environmental and management factors related to mastitis e.g. hygiene measures, but it included specific information on housing and feeding with potential impact on lameness.

Table 2 shows the feature category importance for different diseases. The impact of the breeding values is rather small whereas environmental factors like feeding or housing show a high impact.

In Figure 2 precision and recall for the three different methods (logistic regression, random forest and XGBoost) and two datasets are presented. More complex methods perform better, especially when using the restricted dataset.

Conclusion and next steps

The studies on disease prediction using the “Efficient Cow” dataset showed promising potential of disease risk prediction using integrated datasets. The challenge remains the collection and integration of a data set with such diverse features.

In the next step the different approaches and methods will be applied on an even more complex and integrated dataset including additional data from daily milking and animal based sensors. The aim of the continued research is the development of a data-driven decision support tool providingearly warnings and enabling interventions before diseases fully emerge.

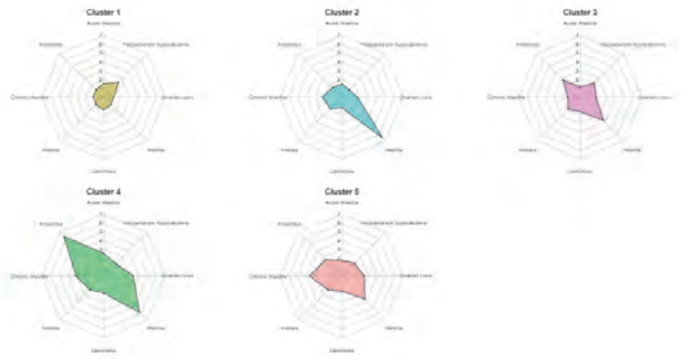


Figure 1. Farm-risk-profiles (Matzhold et al. 2021).

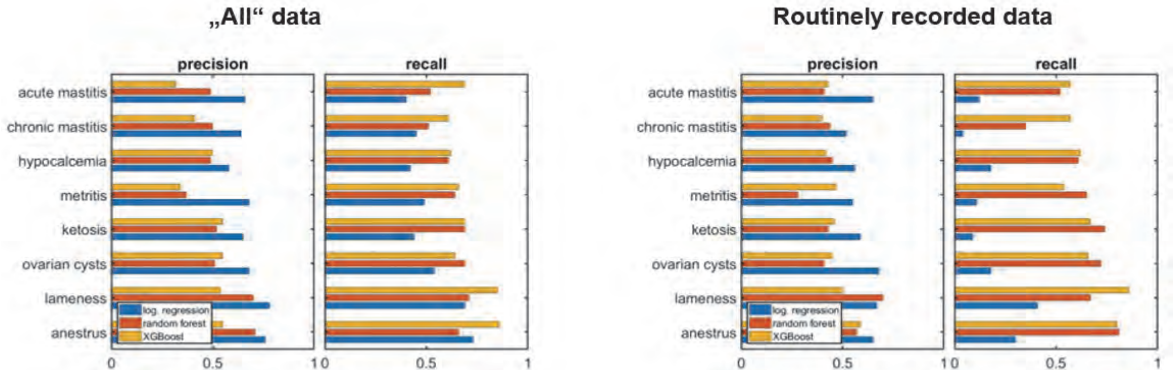


Figure 2: Precision and recall from different methods applied on 2 different datasets (Lasser et al. 2021).

Table 1. Logistic regression results (F1-Score, Precision, Recall) based on the “Efficient Cow” dataset (Lasser et al., 2021).

Disease	F1 Score	Precision	Recall
Anestrus	0,739	0,763	0,729
Lameness	0,737	0,780	0,700
Ovarian cysts	0,616	0,675	0,543
Ketosis	0,521	0,651	0,437
Metritis	0,549	0,677	0,490
Hypercalcemia	0,482	0,576	0,420
Chronic mastitis	0,514	0,635	0,445
Acute mastitis	0,479	0,656	0,395

Table 2. Feature category importance (Lasser et al., 2021).

Feature category	Anestrus	Lameness	Ovarian cysts	Ketosis	Metritis	Periparturient hypocalcemia	Chronic mastitis	Acute mastitis
Age	4,9	36,7	2,8	4,7	2,2	28,4	7,1	11,3
Breed	1,8	2,6	1,2	1,5	2,4	0,4	0,3	2,5
Breeding values	3.0	2,4	2,7	1,7	2,7	1,9	4.0	2,5
Diagnosis source	16.0	15,6	29,6	11,9	33,9	7,1	18,9	16,3
Environment	14,3	4,9	16,5	10,5	15,6	13,8	19,8	10,1
Feed	27,4	11,8	17,5	22,4	12,7	7,5	18.0	19,9
Housing	11,5	12,9	11,3	19.0	15,6	5.0	7,4	13,6
Husbandry	9,1	5.0	4,5	8,8	7.0	2,7	6,4	6,4
Lactation stage	2,7	1,2	4,7	16,1	1,8	28,9	0,7	1,3
Milk indicators	5,9	2.0	3,6	0,7	2,2	1,4	10,4	12,5
Physical indicators	3,6	4,8	5,7	2,7	3,9	3.0	7,1	3,7

[1] Cumulative permutation feature importance contributions for the eleven feature categories. Values are given in % of the sum of all feature importances for a given disease.

Acknowledgement

This work was conducted within the COMET-Project D4Dairy (Digitalisation, Data integration, Detection and Decision support in Dairying, Project number: 872039) that is supported by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK), the Austrian Federal Ministry for Digital and Economic Affairs (BMDW) and the provinces of Lower Austria and Vienna in the framework of COMET-Competence Centers for Excellent Technologies. The COMET program is handled by the FFG.

References

Egger-Danner, C., B. Fuerst-Waltl, C. Fuerst, L. Gruber, S. Hörtenhuber, A. Köck, M. Ledinek, C. Pfeiffer, F. Steininger, R. Weissensteiner, *et al***.** (2017) Efficient Cow-Analyse und Optimierung der Produktionseffizienz und der Umweltwirkung in der Österreichischen Rinderwirtschaft In: Abschlussbericht. Tech. Report. Wien: Zentrale Arbeitsgemeinschaft Österreichischer Rinderzüchter (ZAR). Available from https://raumberg-gumpenstein.at/ forschung/infothek/downloads/ download.html?path=Forschungsberichte%252F12_2017_ gruber_efficient_cow_ abschlussbericht.pdf [accessed November 3, 2021].

Lasser, J., Matzhold, C., Egger-Danner, C., Fuerst-Waltl, B., Steininger, F., Wittek, T., and Klimek, P. (2021) Integrating diverse data sources to predict disease risk in dairy cattle – a machine learning approach. Journal of Animal Science 99(11), skab294.

Matzhold, C., Lasser, J., Egger-Danner, C., Fuerst-Waltl, B., Wittek, T., Kofler, J., Steininger, F., and Klimek, P. (2021) A systematic approach to analyse the impact of farm-profiles on bovine health. Science Reports 11, 21152. doi: 10.1038/s41598-021-00469-2

Making use of dairy herd improvement records and machine learning to identify best management strategies

D. Warner[1], J. Gunn[1] and D.E. Santschi[1]

Lactanet, 555, boul. des Anciens-Combattants, Sainte-Anne-de-Bellevue, H9X 3R4,
Québec, Canada
Corresponding Author: dwarner@lactanet.ca

Abstract

Dairy herds continuously generate on-farm data that have a high potential for decision making support. However, variability among herds can be high and it can be, therefore, challenging for dairy producers and extension services to interpret farm records and identify best management practices. In this observational study, we used DHI data to identify the main factors explaining the overall herd performance for 71 commercial Holstein herds in New Brunswick, Canada. A machine learning approach was used to generate high-level recommendations as well as detailed herd-specific recommendations. Enrolled farms were equipped with a parlour (n = 31), pipeline (24) or automatic (16) milking system. Herd performance records from September to December 2020 were assembled from a DHI database and included 71 indicators related to production, reproduction, health, and longevity. Herds were segmented based on their overall herd performance using hierarchical clustering based on principal components. The principal components were used primarily as pre-processing step to de-noise the data and to balance the influence of the similar DHI records. In-dept analysis of clusters was conducted using decision tree induction with the aim to generate an interpretable on-farm decision making tool. Three herd clusters were identified and consisted of low overall herd performance across all DHI indicators (Cluster 3; n = 17), a medium overall herd performance but high longevity (Cluster 1; n = 36), and high overall herd performance but with a concomitant low longevity (Cluster 2; n = 18). Decision tree induction further allowed to identify the most important DHI indicators explaining the adherence of a herd to one of the three clusters. This ultimately allowed to establish high-level recommendations and visualize which performance indicators a herd might want to focus on to improve overall herd performance. For instance, low performing herds were observed to be mainly driven by a low reproduction performance. In addition, single predictions of the decision tree algorithm were fitted with a local interpretable model. This latter approach allowed to add interpretability to the decision tree model and generate dynamic herd-specific recommendations for each farm. In conclusion, the results suggest that mining DHI data can give a valuable insight into best herd management practices and can be used to highlight opportunities for improvements.

Keywords: Decision support, DHI, artificial intelligence, decision tree.

Introduction

Dairy herds continuously generate on-farm data that have a high potential for decision making support. However, variability among herds can be high and it can be, therefore, challenging for dairy producers to interpret farm records and identify best management practices. In addition, on-farm data from dairy herd improvement (DHI) records are typically complex, incomplete, variable over time, and correlated to each other. With such datasets, traditional inferential statistical methods often perform poorly (Heald *et al*., 2000). A machine learning approach might be more suitable for DHI data as these types of algorithms provide more flexibility with incomplete data and generally handle better interactions among predictors and non-linear relationships (Abbot, 2014). Unsupervised machine learning algorithms such as clustering are typically used to find hidden patterns in complex and large datasets, whereas supervised machine learning algorithms are typically used for generalized predictions. In addition, machine learning is particularly suitable for field applications as it can be implemented and automated for field applications (Dominiak and Kristensen, 2017). As such, a combined unsupervised and supervised machine learning approach was used in the present study to find groupings in data originating from the dairy herd population in New Brunswick, Canada, and to identify management practices of high performing dairy herds as opposed to other dairy herds for decision making support to improve herd performance and profitability.

Material and methods

Participating herds

A total of 89 dairy herds across New Brunswick for which DHI records are routinely collected were initially included in the study. Only herds predominantly composed of Holstein cows (71 herds) were considered for further analysis as herd management was likely influenced to a large degree by the breed. Among these, 31 herds had a parlour, 24 herds had a pipeline, and 16 herds had an automatic milking system. Each participant received free access to the Lactanet service for transition management (Transition Cow Index; Nordlund, 2006) and subclinical ketosis screening (Foss, 2009) during the entire length of the study period which were used as input data in the prediction model.

Data collection

Dairy herd improvement data.

Test day records were collected and included 12-month average data on herd demography, production, longevity, and reproduction based on the last test records registered between September and December 2020. Data was cleaned to remove implausible values and outliers by removing the 1st and 99th percentile. Originally 71 variables were considered for data analysis. Variables missing at random were imputed using a random forest approach as described in van Buuren (2012), which generally handles well complex and inter correlated data in the presence of missing cases (Tang and Ishwaran, 2017) as it the case for DHI data. Data not missing at random (bodyweight, milk urea nitrogen records) were omitted from the analysis.

Survey data

Management practices were collected through an online survey administered by Lactanet technicians and advisors. The survey was previously developed and applied to collect data on approximately 2300 Quebec dairy herds with questions pertaining to the building, housing, feed bunk and feeds, bedding, cleaning, milking system, footbath, hoof trimming, exercise and pasture, calving, and drying off management (Lactanet, 2021). Among the 71 Holstein herds in New Brunswick, 42 completed the survey and were thus considered for analysis. Amongst these, 23 herds had a parlour, 22 herds a pipeline, and 12 herds an automatic milking system. The survey was completed between March 2020 and April 2021.

Data analysis

To identify the best management practices, herds were clustered into similar groups based on the overall herd performance using their respective test day DHI records by means of agglomerative hierarchical clustering on principal clustering (HCPC) using the FactoMineR package in R (version 3.5.0; R Foundation for Statistical Computing, Vienna, Austria) of Le *et al.* (2018). Computation of the principal components was thereby done as pre-processing step to de-noise the data and to balance the influence of several groups of variables.

Differences in the DHI records among the identified clusters were first evaluated through a mixed effect linear regression with cluster considered fixed and milking type considered random using the lmerTest package (Kuznetsova *et al*., 2017).

The cluster output was investigated in more detail through a decision tree approach with the aim to predict the adherence to one of the three groups with the DHI indicators and identify the most meaningful indicators explaining the difference between the high, medium, and low performant herds. Decision tree induction was conducted through a CART classification tree as described in Breiman *et al*. (1984) using the caret modelling package workflow of Kuhn (2008). This approach allows to compute a tree-structured classification containing a collection of decision rules (represented by the tree branches) and cluster predictions (represented by the terminal node of a branch). Its simple and interpretable structure as well as its ability to compute decision rules in the presence of missing values using surrogate variables (i.e., substitutes for the primary splitter of a node when the primary splitter is missing; Rokach and Maimon, 2005) makes it an interesting tool for future field applications.

Prior to applying a decision tree algorithm, collinearity was checked to remove variables with a Spearman correlation coefficient above 0.60 as multicollinearity could lead to over-fitting, resulting in 19 final variables. The optimal model hyperparameters for the tree size (complexity parameter) were evaluated through a random search using a 5-fold cross-validation.

The global variable importance scores, that is the variables contributing most to the model development, were extracted based on the sum of the reduction in the loss function (information gain) attributed to each variable at each split. The model response was illustrated through an alluvial diagram for the variables with the higher global importance score. The variables were thereby normalized to a mean of 0 and a standard deviation of 1. Additional model agnostics based on the LIME approach (Local interpretable model-agnostic explanations; Ribeiro *et al*., 2016) were run to extract local variable importance scores, which allow to interpret the model outcome for each participating herd.

The survey data were analysed using a Fisher's test on the cluster output. The Fisher's test was preferred over the more commonly used chi-squared test due to the small sample number and failing to comply with the strict requirements of the latter (e.g., expecting a frequency never smaller than 1 and a frequency of 5 or more for at least 80% of the output cells). A list of promising management practices known to affect herd performance has been elaborated and presented to the participating producers. However, as the findings were statistically inconclusive due to the low response rate and large variability observed among the participating herds leading to an overall non-significant P-value (> 0.10), results from the survey are not shown here.

Each participant received a customize report with benchmarks for DHI records previously identified using decision tree induction as well as for some relevant management practices via a parameterized R markdown report (version 1.1.; Allaire *et al*., 2019) and were directly sent to the respective participant from the RStudio user interface via the utility package mailR (version 0.4.1; Premraj, 2015).

Results and discussion

Dairy herd improvement data

Three cluster were identified based on the herd DHI records and labelled as high, medium, and low overall herd performance. A detailed cluster description is shown in Table 1 for variables that differed significantly ($P \leq 0.05$) among clusters. The herd size did not differ significantly among the clusters ($P = 0.076$) but was somewhat higher for the Medium performance cluster due to a higher number of herds equipped with a milking parlour. The Low performance cluster did not contain any herds equipped with an automatic milking system. However, herds pertaining to the High performance cluster were not driven mainly by a specific milking system.

The cluster analysis suggested that the High performing herds were also the herds with the highest milk performance, milk value (mainly based on milk yields and components) and genetic potential. The somewhat higher genetic potential for energy-corrected milk for the Medium performant herds implies that some herds in this cluster do not exploit

Table 1. Cluster description for overall herd performance.

	Cluster		
Variable	**High**	**Medium**	**Low**
Number of herds	18	36	17
Herd size (median)	75.2	97.2	75.0
Milking system (number of herds)			
Automatic	6	10	0
Parlour	6	17	8
Pipeline (tie stall)	6	6	9
Production[1]			
kg ECM	10,996	9,909	8,468
kg fat	442	397	339
kg protein	361	328	277
Milk value (CAN$)	8,032	7,188	6,127
Genetic index for ECM	414	440	255
Reproduction and health[1]			
Transition Cow Index[2]	426	260	−173
Calving interval	405	403	451
Days to first breeding	82	81	100
Age at first calving	25.6	25	30
Days dry	67	63	71
% cows with SCC > 200,000	12.4	14.8	18.4
Longevity[1]			
% involuntary culled	29	16	23
% cows dead	4.7	2.8	5.3
% cows left	44	30	35
% cows left at 60 DIM	8.3	5.2	7.3
% cows left for reproduction	8.8	4.6	5
% cows left for feet problems	3.7	1.9	4.2

their genetic potential to the fullest as suggested by their lower milk performance and milk value. In addition, the High performing had the best transition cow management and a high reproduction performance. This further highlights the importance to pay attention to the non-producing cows and heifers within a herd. Nonetheless, these herds had a high involuntary culling rate and deficiencies in early lactation, in particular compared to the Medium performance cluster indicating a potential for improvement.

Using a decision tree approach, the most important variables were identified that explain the adherence of a herd to either the High, the Medium and Low performing herd cluster (Figure 2). The top performing herds could be predicted via a low calving interval, low age at first calving, high involuntary culling rate, high energy-corrected milk yield, high turnover rate and percentage of cows left, and an excellent transition cow management. The decision tree suggested that the contribution of the genetical potential was low overall suggesting that some lower performing herds might not fully exploit their genetic potential.

Using a model-agnostic approach, the model predictions were investigated in more detail. For a randomly selected herd (herd 61 in Figure 2), we noticed a high calving interval, suboptimal transition management and low turnover rate which were in line with other Low performing herds. However, its genetic potential was overall within the top 50% herds and thus considerably higher than that of its peers, suggesting that herd 61 does not fully exploit its high genetic potential. This approach allows therefore to investigate the decision rules for each individual herd and can help with decision

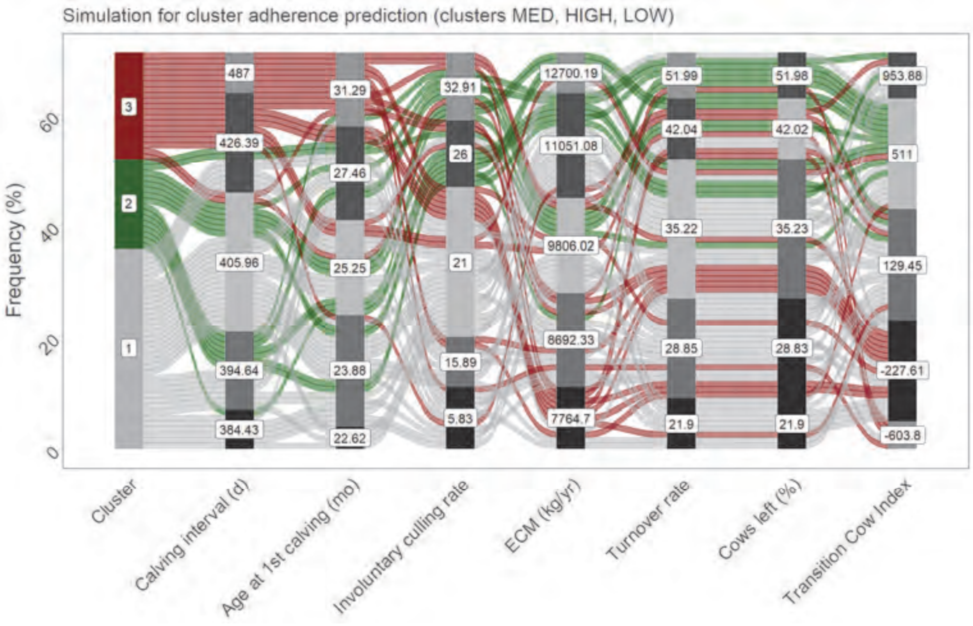


Figure 1. Model response, as shown by an alluvial diagram with predicted cluster adherence as affected by the most important predictors ranked by their variable importance score. Nodes on the left denote the Medium (1), High (2), and Low (3) performance cluster; nodes for predictors with respective means were fixed to five nodes based on the frequency distribution (y-axis).

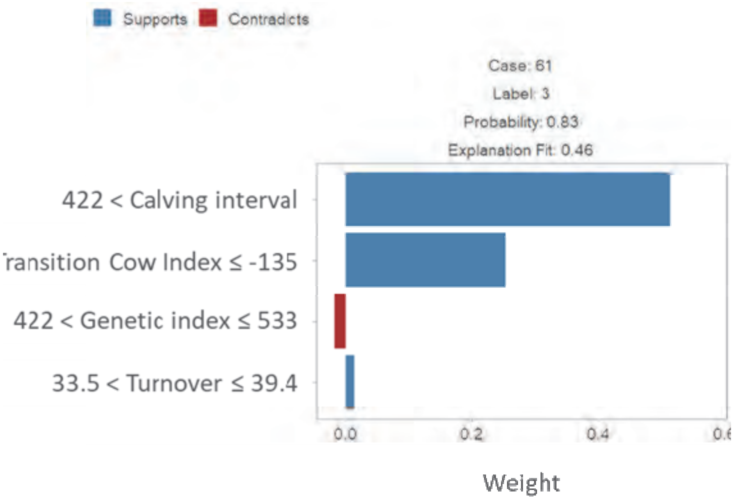


Figure 2. Contribution of variables of a randomly selected herd pertaining to the Low performing cluster (Cluster 3) based on a model agnostic approach.

making by highlighting possible similarities or discrepancies among herds within the same cluster and by pinpointing to potential deficiencies for a producer to focus on.

Conclusion

The findings presented here suggest that mining DHI data can give a valuable insight into best herd management practices and can be used to highlight opportunities for improvements. DHI records have a high potential for decision making support. Due to the type of records at hand, a more flexible approach using machine learning is recommended to deal with the large data volume and variety. Clustering allowed to identify 3 main herd clusters based on the overall herd performance using DHI data. High performing herds differentiated themselves from low performing herds mainly by adopting an optimal transition management, overall good reproduction performance and high attention given to the non-producing animals despite a potential for improvement in early lactation management. Decision tree induction further allowed to identify the most important DHI indicators explaining the adherence of a herd to one of the three clusters. This ultimately allowed to establish high-level recommendations and visualize which performance indicators a herd might want to focus on to improve overall herd performance. A local interpretable model fitted to single decision tree predictions allowed to add interpretability to the decision tree model and help with decision making by highlighting specific possible deficiencies within a herd.

Acknowledgement

This project was funded by Agriculture and Agri-Food Canada (Ottawa, ON, Canada) and Milk 2020 (Sussex, NB, Canada) through the Canadian Agricultural Partnership joint research program.

References

Abbott, D. 2014. Applied predictive analytics: Principles and techniques for the professional data analyst. Wiley and Sons, Indianapolis, IN.

Allaire, J.J., J. Horner, Y. Xie, V. Marti and N. Porte. 2019. markdown: Render Markdown with the C Library 'Sundown'. R package version 1.1. https://CRAN.R-project.org/package=markdown.

Breiman, L., J.H. Friedman, R.A. Olsen and C.J. Stone. 1984. Classification and regression trees. Wadsworth and Brooks, Monterey, CA.

Dominiak, K.N. and A.R. Kristensen. 2017. Prioritizing alarms from sensor-based detection models in livestock production - A review on model performance and alarm reducing methods. Computers and Electronics in Agriculture 133: 46–67. https://doi.org/10.1016/j.compag.2016.12.008.

Foss. 2009. Foss Application Note 35. MilkoScan FT+ Ketosis Calibrations. Foss, Hillerød, Denmark.

Heald, C.W., T. Kim, W.M. Sischo, J.B. Cooper and D.R. Wolfgang. 2000. A computerized mastitis decision aid using farm-based records: an artificial neural network approach. Journal of Dairy Science 83(4): 711–720. https://doi.org/10.3168/jds.S0022-0302(00)74933-2.

Kuhn, M. 2021. caret: Classification and Regression Training. R package version 6.0-88. https://CRAN.R-project.org/package=caret.

Kuznetsova, A., P. B. Brockhoff and R.H.B. Christensen. 2017. lmerTest Package: Tests in linear mixed effects models. Journal of Statistical Software 82(13): 1–26. doi: 10.18637/jss.v082.i13.

Lactanet, 2021. Portrait des fermes laitières québécoises en matière de logement et de gestion des animaux. Lactanet: https://lactanet.ca/wp-content/uploads/2022/01/brochure_portrait-ferme-qc_021v11.pdf (accessed on May 2022).

Le, S., J. Josse and F. Husson. 2018. FactoMineR: An R package for multivariate analysis. Journal of Statistical Software 25: 10.18637/jss.v025.i01.

Nordlund, K., 2006. Transition Cow Index. In: Proceedings of the 39th Annual Conference of American Association Bovine Practitioners, St. Paul, MN, USA, 20–24 September 2006; American Association of Bovine Practitioners, Ashland, OH, USA, p 139–143.

Premraj, R. 2015. mailR: A Utility to Send Emails from R. R package version 0.4.1. https://CRAN.R-project.org/package=mailR.

Ribeiro, M.T., S. Singh and C. Guestrin. 2016. Why should I trust you? Explaining the predictions of any classifier. arXiv:1602.04938 [cs.LG].

Rokach, L. and O. Maimon. 2005. Decision trees. In: Data Mining and Knowledge Discovery Handbook, Maimon, O. and L. Rokach (Editors), Springer, Boston, MA, p 165–192.

Tang, F. and H. Ishwaran. 2017. Random forest missing data algorithms. Statistical Analysis and Data Mining 10 : 363–377. https://doi.org/10.1002/sam.11348.

van Buuren, S. 2012. Flexible imputation of missing data. CRC Chapman and Hall, Boca Raton, FL.

Prediction of grass-based diet from indirect traits using milk MIR-based predictors to assess the feeding typology of farms

H. Soyeurt[1], C. Gerards[1], C. Nickmilder[1], S. Franceshini[1], F. Dehareng[3], D. Veselko[4], C. Bertozzi[5], J. Bindelle[1], N. Gengler[1], A. Marvuglia[6], A. Bayram[6,7] and A. Tedde[1,2]

[1]TERRA, Gembloux Agro-Bio Tech, University of Liège, Passage des déportés 2, 5030 Gembloux, Belgium
Corresponding Author: hsoyeurt@uliege.be
[2]FNRS, Rue d'Egmont 5, 1000 Brussels, Belgium
[3]Walloon Agricultural Research Centre, Valorisation of Agricultural Products, Chaussée de Namur 24, 5030 Gembloux, Belgium
[4]Comité du lait, Route de Herve 104, 4651 Battice, Belgium
[5]Walloon Breeding Association, Rue des Champs Elysées 4,5590 Ciney, Belgium
[6]Luxembourg Institute of Science and Technology (LIST), Rue du Brill 41,4422 Belvaux, Luxembourg
[7]Computational Sciences, Faculty of Science, Technology and Medicine, University of Luxembourg, Avenue de l'Université 2, 4365 Esch-sur-Alzette, Luxembourg

Abstract

Grassland provides access to a local and low-cost food resource usable by livestock. Today, for some brands of dairy products, one of the specifications asked the farmers is to put their cows on pasture for a minimum period or feed their cows mainly with a grass-based diet. But this constraint is rarely verified. However, this kind of feeding has a fingerprint in the milk. So, this work aims to develop a method predicting indirectly the level of grass in the cows' diet using features estimated from milk mid-infrared (MIR) spectrometry. More than 3 million records were collected between 2011 and 2021 on 2,449 farms. Those included the fat, protein, lactose, and urea contents estimated by the spectrometers and 34 MIR-based predictors reflecting the milk fatty acid composition, the protein fraction, the minerals, and the lactoferrin content. As no grazing calendars and detailed feed composition were available at large scale, the innovative aspect of this work will consist of estimating the grass-based diet using a trait defined from the month of analysis of bulk milk. Indeed, in Wallonia, nearly all herds are on pasture between the beginning of sprint and the beginning of autumn. The data were collected between 2019 and 2020. The calibration set used 80% of the farms chosen randomly (345,223 records). The remaining ones were used to validate the model (85.069 records). The direct prediction of test month from a Partial Least Square discriminant analysis (PLS-DA) was poor (0.40). However, a hierarchical clustering applied to those obtained predictions revealed a GRASS group composed of predictions done in June, July, August, and September. So, a second PLS-DA was realized to discriminate the GRASS group. A better accuracy was good (0.88). Then, the probability of belonging to the GRASS modality was used to observe the feeding typology of the farm. The evolution of this probability through the year was the one expected with an increasing value from April and a constant decrease after summer. The interval between this increase and decrease could be used to count the number of days spent by cows on pasture and the intensity of this change can be studied discriminate different feeding strategy. In conclusion, using the MIR analysis done for the milk payment by dairy companies, it is feasible to detect the presence of grass in the cow diet and estimate potentially the number of days spent by the cows on pasture during a year.

Keywords: Grass, milk, mid-infrared.

Introduction

In the Southern part of Belgium, grassland represents a large part of the useful agricultural surface. Permanent grasslands represent 42.1% of this surface (309,180 ha) and temporary grasslands represent 5% (State of the Walloon Environment, 2020).

Grassland has an important place in our landscape and fulfils many roles. But grassland provides also access to a lower cost feed resource, requiring very little supplementation for ruminants (Kolver, 2003). Therefore, producing milk from pasture makes sense. Many specifications for milk and protected designation of origin products require farmers to have a minimum period of grazing for their cows. For example, in France, CANDIA recommends an average of 150 days of grazing per year and for at least 6 hours per day. The “Lait de pâturage” label is based on the same rules but also requires a minimum grazing area of 10 ares per animal. Finally, the “Grand Pâturage” milk requires its members to have cows that have access to at least 180 days of pasture per year. In Belgium, the “MARGUERITE HAPPY COW” chain advocates feeding cows with a minimum of 70% grass in the ration and at least 180 days of grazing, for a cattle load of 4 animals per ha (Servais, 2015).

However, even though these specifications exist, there is no real verification of these conditions. Thus, recently, the Chronopâture tool was developed to automatically count the number of days that the herd spends grazing, using a GPS collar. Unfortunately, such equipment can hardly be developed on a large scale, as it depends on the choice of the breeder. An alternative could be to observe more specifically the variations of milk composition according to the feed. For example, Coppa *et al* (2021) demonstrated that it was possible to predict the composition of the cow’s ration and to identify PDO practices using mid-infrared spectrum (MIR) analysis of milk, but they did not attempt to predict the number of days on pasture. This will be the aim of this work with the innovative aspect to use an indirect reference to predict the grass-based diet.

Materials and methods

The data come from the FuturoSpectre agreement linking Gembloux Agro-Bio Tech (University of Liège, Gembloux, Belgium), the Walloon Agricultural Research Centre (CRA-W, Gembloux, Belgium), the milk laboratory ‘Comité du Lait’ (Battice, Belgium), and the Walloon Breeding Association (AWé, Ciney, Belgium). Milk samples were collected in the bulk tank of 2,868 Walloon dairy farms between 2019 and 2020. The dataset contained 430,292 observations. All milk spectral data were generated by MilkoScan FT6000 MIR spectrometers (Foss Electric A/S, Hillerød, Denmark). Equations were applied on the recorded spectral data to estimate 38 traits related to the milk fatty acid (FA) profile, composition of mineral and protein as well as the content of lactoferrin and beta-hydroxy butyrate. All equations have a prediction R^2 higher than 65%. The FA contents initially predicted by MIR in g per dl of milk were converted into g per 100 g of fat using the fat content provided by the spectrometer.

The prediction of grass-based diet is not directly possible as we do not have data on grazing management. However, in the Walloon Region of Belgium, there is grazing between the beginning of spring and the beginning of autumn. So, the hypothesis formulated in this work is that there is an indirect link between the month of the milk analysis and the presence on pasture. This month of test was so predicted using Partial Least Square (PLS) regression. To allow an external validation of the developed prediction equations, the initial dataset was divided into two parts: 80% training set and 20% validation set. The former includes 2,287 farms (345,223 records) while the latter includes 581 farms (85.069 records). To ensure the data independence, the partition was performed with the constraint that the same farm could not be found in both sets. Based on the obtained predictions for the month, a hierarchical clustering was performed to highlight a "GRASS" group and a "NOGRASS" group. Then a second prediction model was performed to predict the GRASS modality. The prediction quality of the model was assessed by calculating the accuracy.

Results and discussion

The accuracy obtained for the prediction of the month of test based on the used features was poor and equal to 0.394 ± 0.002. The accuracy of the external validation was of 0.398. However, a strong confusion existed between some months (Figure 1). Based on the dendrogram created after the hierarchical clustering done on the predicted values, it appeared clearly 2 groups. The first group that covered June, July, August, and September can be attributed to predicted of grass-based diet (i.e., GRASS modality). Therefore, a second PLS regression was built to predict the GRASS modality. The obtained prediction was, as expected, strongly better with an accuracy of 0.874 ± 0.002. The validation accuracy was similar (0.876). This confirms the findings of Frizzarin *et al*. (2021) and Coppa *et al*. (2021) who mentioned that PLS-DA has very good ability to predict cow diet from the MIR spectrum.

The feeding fingerprint explains this capacity to discriminate the grass-based diet in milk spectral data. According to Chilliard *et al*. (2000), fresh grass is the richest food in C18:3 fatty acids. Therefore, its ingestion by cows will confer a particular composition to the milk. A literature review done by Elgersma *et al*. (2006) revealed

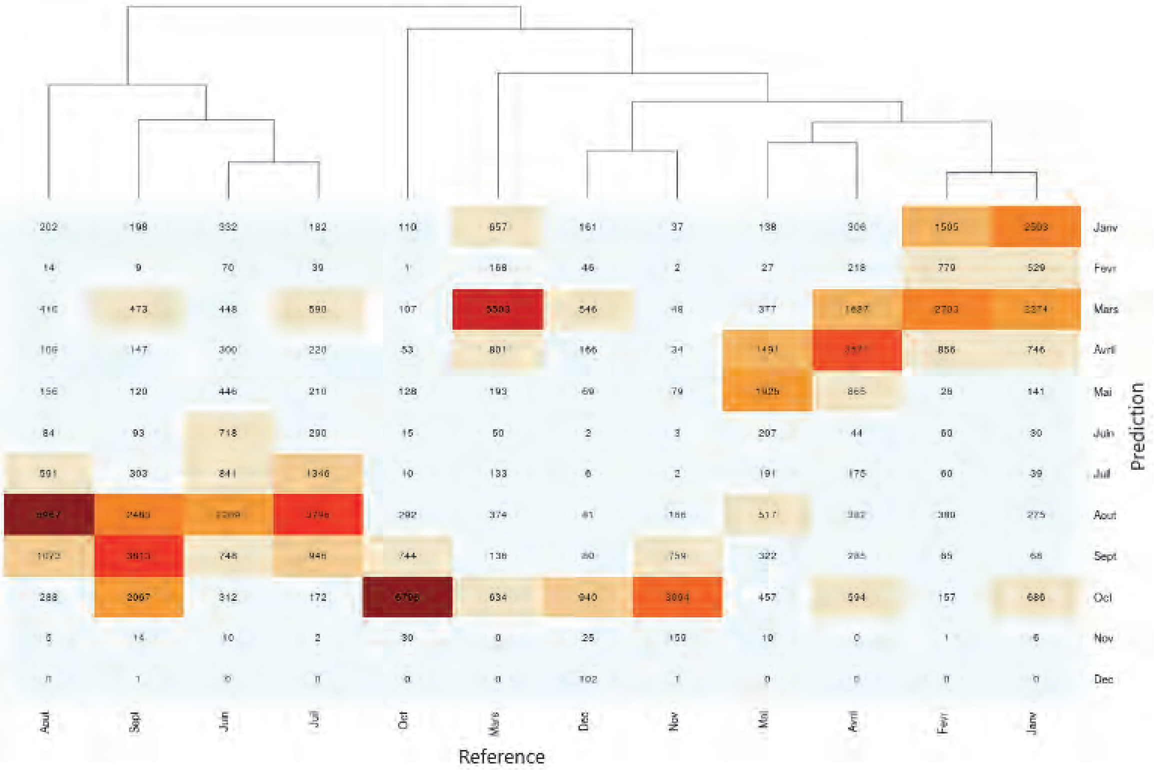


Figure 1. Dendrogram and confusion matrix between predicted and observed months.

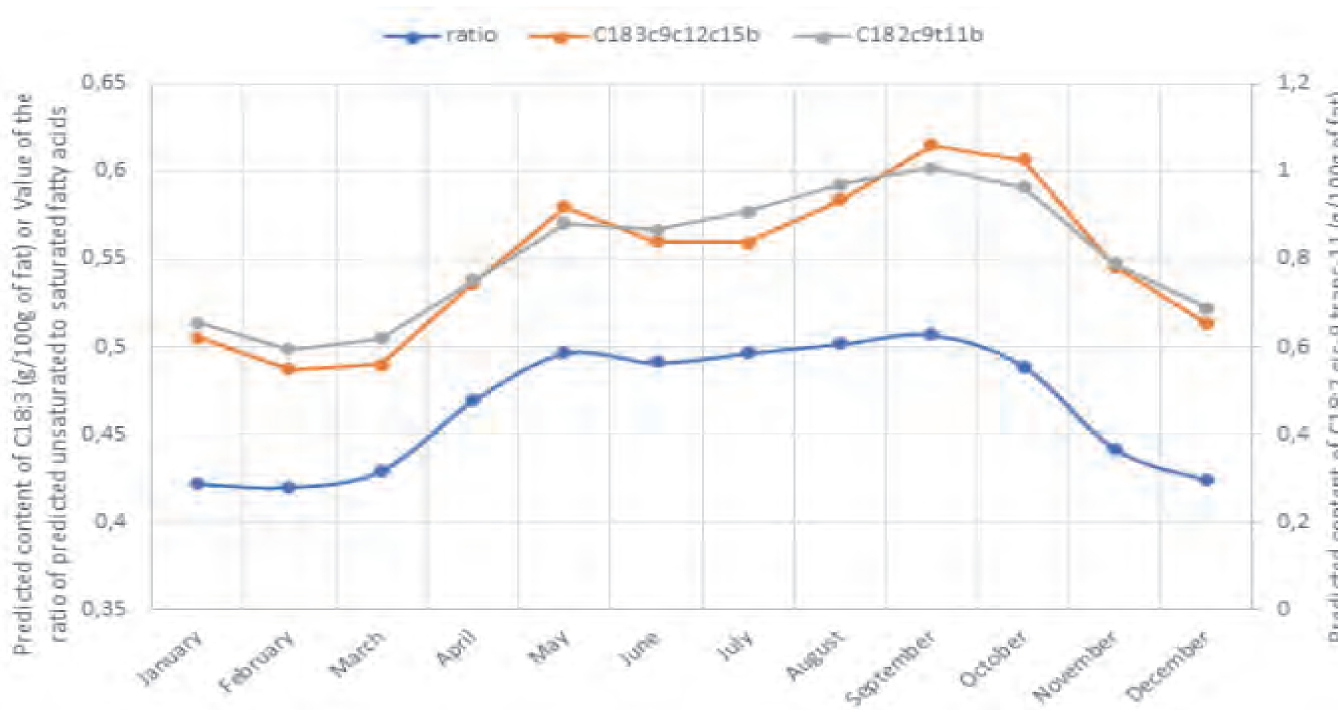


Figure 2. Yearly evolution of the ratio of predicted unsaturated to saturated fatty acids and the predicted content of C1 cis-9,trans-11 and C18:3 cis-9, cis-12,cis-15 in milk fat.

that the milk of cows fed with fresh green forage has a higher ratio of unsaturated to saturated fatty acids and contains a higher proportion of polyunsaturated fatty acids and conjugated fatty acids (CLA) than the milk of cows fed with a total mixed ration. Even if the prediction accuracy for those FA traits by FT-MIR spectroscopy is moderate, the same trends were observed from our predicted FA values during the grazing period (Figure 2). Moreover, Frelich *et al*. (2012) observed an increase in long-chain FAs (LCFA) and a decrease in short- (SCFA) and medium-chain FAs (MCFA) in the milk fat when the feeding of cows is based on fresh grass. Again, this was also confirmed by our predicted FA values (Figure 3).

The presence of grass-based diet is not a binary variable. Indeed, there is a feeding transition between winter and summer diets. The composition of the milk will therefore evolve gradually. Therefore, the use of a binary prediction does not make sense.

Thus, it is much more relevant to look at the evolution of the prediction probability related to the presence of grass-based diet. Figure 4 represents the distribution of the GRASS PLS probability following the week of test in 2019. The annual evolution of these probability values has an expected trend with an increasing probability from April, which represents the passage to pasture, and then a decrease when the winter arrived, signifying the return of cows to the barn and a total mixed ration as feeding. So, the interval between the increase and decrease of the probability of belonging to the GRASS modality could be used to calculate the number of days spent by dairy cows on pasture.

8:2

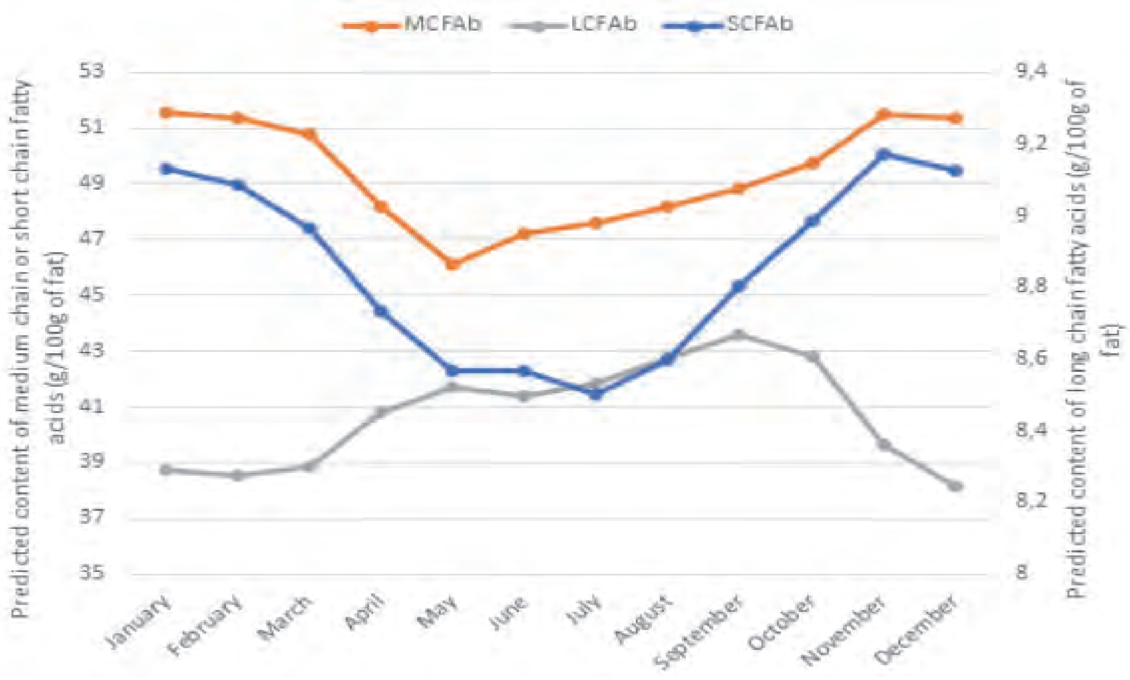


Figure 3. Yearly evolution of the predicted content of short, medium, and long chain fatty acids in milk fat.

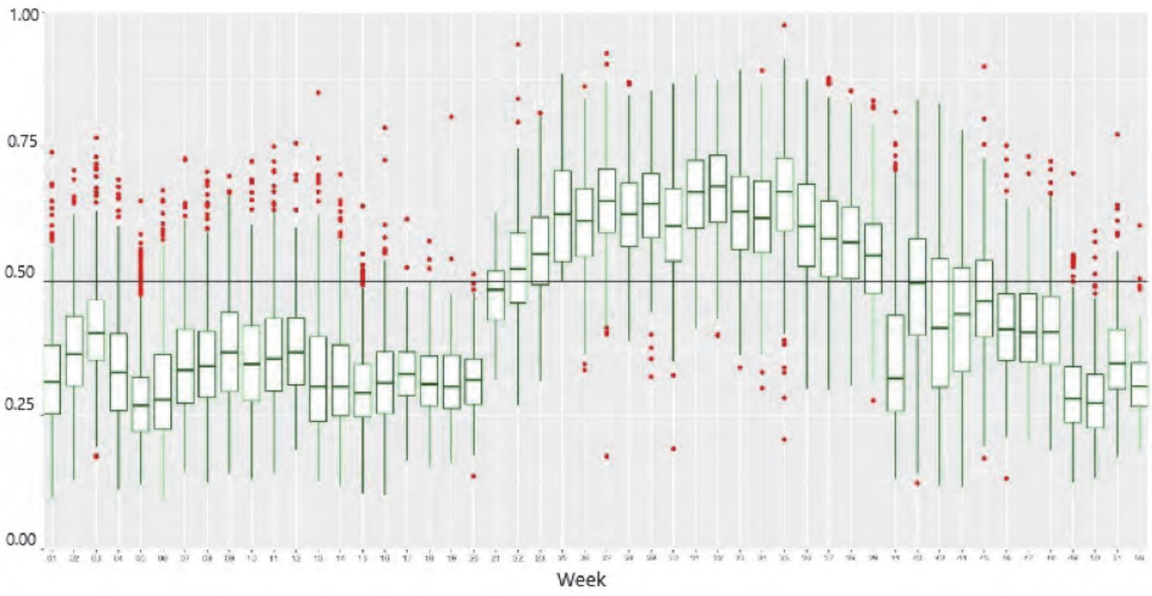


Figure 4. Boxplot of probabilities of belonging to the GRASS group in 2019 according to weeks (N = 101,442).

Conclusion

The objective of this work was to contribute to the development of a tool for counting days on pasture for dairy cows on a given farm automatically. To achieve this objective, model was built to predict the presence of grass-based diet from the spectral data provided by the analysis of milk collected in the bulk tank. As no grazing calendar was available, this innovative aspect of this study was to propose the development of a model predicting an indirect trait, the test month. The month confusion during the prediction revealed the possibility to create a GRASS group. So, the second model was built to predict this group with a good accuracy. The evolution of the PLS probability had the expected trend and suggest the possibility to count the number of days spent on pasture. This measurement as based on the milk mid-infrared spectrometry presents the advantage to be cheap and available for all farms delivered their milk to a dairy company. This calculation is also interesting as the frequency of data acquisition is high. Indeed, the bulk tank milk samples are analysed every 1 or 4 days in each farms delivered milk to the dairy company.

Acknowledgment

This work was supported by the Luxembourg National Research Fund (INTER/FNRS/18/12987586/SimBa) and by the Fonds de la Recherche Scientifique-FNRS under Grant n° T.0221.19.

References

Chilliard, Y., Ferlay, A., Mansbridge, R.M., Doreau, M., 2000. Ruminant milk fat plasticity:nutritional control of saturated, polyunsaturated, trans and conjugated fatty acids. Ann. Zootech. 49, 181–205.

Coppa M., Revello-Chion A., Giaccone D., Ferlay A., Tabacco E. and Borreani G., 2014. Comparison of near and medium infrared spectroscopy to predict fatty acid composition on fresh and thawed milk. Food Chem. 150, 49–57.

Frizzarin M., O'Callaghan T.F., Murphy T.B., Hennessy D. and Casa A., 2021. Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets. J. Dairy Sci. 104, 12394-12402.

Elgersma, A., Tamminga, S., Ellen, G., 2006. Modifying milk composition through forage. Anim. Feed Sci. Technol. 131, 207–225.

Frelich, J., Šlachta, M., Hanuš, O., Špièka, J., Samková, E., Zapletal, P., 2012. Seasonal variation in fatty acid composition of cow milk in relation to the feeding system. Animal Science Papers and Reports 30, 219-229.

Kolver, E.S., 2003. Nutritional limitations to increased production on pasture-based systems. Proc. Nutr. Soc. 62, 291–300.

Servais L. (Awé asbl), 2015. Marguerite happy cow une nouvelle filière laitière durable. Wallonie Elev. - Mai 2015 51–52.

Applying fatty acid profiles from bulk tank milk on farm for decision-making support

D. Warner[1], R. Lacroix[1], R. Gervais[2] and D.E. Santschi[1]

[1]*Lactanet, 555, boul. des Anciens-Combattants, Sainte-Anne-de-Bellevue, H9X 3R4, Québec, Canada*
[2]*Département des sciences animales, Université Laval, 2425, rue de l'Agriculture, Québec, G1V 0A6, Québec, Canada*
Corresponding Author: profilab@lactanet.ca; dsantschi@lactanet.ca

Abstract

Routine analysis of fatty acids (FA) in commercial bulk tank samples through Fourier-transform infrared (FTIR) spectroscopy is increasingly being used as a dairy management tool to assess feeding and farm management practices. Bulk tank FTIR milk FA profiles from 3,200 dairy herds in Quebec, Canada, over 3 years (April 2019–March 2022; 1.9M samples) were assembled in a national database (Lactanet, Canada) and matched with the nearest test day records. The emphasis was put on the de novo, mixed and preformed FA groups as they originate from synthesis in the mammary gland (*de novo*), feed intake or body fat mobilization (preformed), or both (mixed). Specifically, *de novo* FA were associated with improved fat and protein content in bulk tank milk ($r_{corr} = 0.81–0.82$). Seasonal changes were distinctive with decreased *de novo* and mixed FA and increased preformed FA during the summer months. Increasing the proportion of corn silage in the ration up to 50% was associated with increased *de novo* and decreased preformed FA (based on a subset of 513 Holstein herds with available feed records). The margin over feed costs (MOFC; in $ per kg of hl milk or per kg of milk fat) was greater with increased *de novo* FA and smaller with increased preformed FA, whereas no association was observed between the milk FA profile and MOFC per cow (based on a subset of 95 Holstein herds with available profitability records). Overall, with increasing herd performance (kg of milk fat/cow and day) *de novo* FA increased, and preformed FA decreased. This trend was inversed for high performing herds (≥ 1.70 kg milk fat/cow and day), which further emphasizes the need to compare herds with specific milk FA profiles to their appropriate peers. However, variation of milk FA profiles was high within cohorts of herds with similar milk performance, thus highlighting the importance for using herd-specific benchmarks in relation to historical farm data and strategic goals for a herd (internal benchmarking). An interactive dashboard application was developed to visualize milk FA profiles including benchmarks for the user to interpret their herd performance against their peers and identify best management practices. To facilitate early detection and decision making on farm, current research efforts include the detection of anomalies and changes in the overall trend of milk FA as well as the identification of their potential causes (diagnostic) through artificial intelligence.

Keywords: Decision support, fatty acid, mid-infrared spectroscopy, benchmarking.

Introduction

Fat is the most variable component of milk. Its concentration and its fatty acid (FA) profile vary according to several factors including the animal (parity, lactation stage, breed), the environment (season, management system) and nutrition (forage and grain type and proportion, amount, and composition of dairy fat supplements). Traditionally analyzed by gas chromatography, a laborious and costly technique, advances in technology over the last decade allow measuring major milk fatty acids routinely and rapidly by mid-infrared (MIR) spectroscopy. This technique offers the opportunity to monitor milk FA in bulk tank or individual cow samples alongside and, thus, at the same frequency as analyzing for major components (e.g., fat, protein, lactose, milk urea nitrogen, somatic cell counts). It is therefore possible to obtain the complete composition of a milk sample, including the FA profile with the main groups of FA. Management and dietary practicing that modulate cow behaviour and rumen conditions were previously found to be associated with changes in the FA synthetized *de novo* from rumen precursors based on northeastern US commercial dairy farms (44 and 39 farms, respectively; Woolpert *et al.*, 2016, 2017). Analyzing these FA routinely in bulk tank samples collected every other day therefore open the possibility to better monitor herd health and performance, and to better understand and monitor herd milk components and overall herd performance. Our main objective was to better understand herd milk FA profiles through benchmarking and the development of a decision-making tool to support producers in their daily management decisions.

Milk fatty acids origin

Milk FA can be divided in 3 main groups, according to their biological origin. These groups include de *novo* FA (short chain FA synthesized in the mammary gland from rumen precursors); Preformed FA (Long chain FA derived from dietary FA transferred into milk and FA mobilized from adipose tissue); and Mixed FA (Medium chain FA from mixed origin; Figure 1). Accordingly, changes in nutrition, management or metabolism of the cows and the herd will be reflected as changes in milk fatty acid profile.

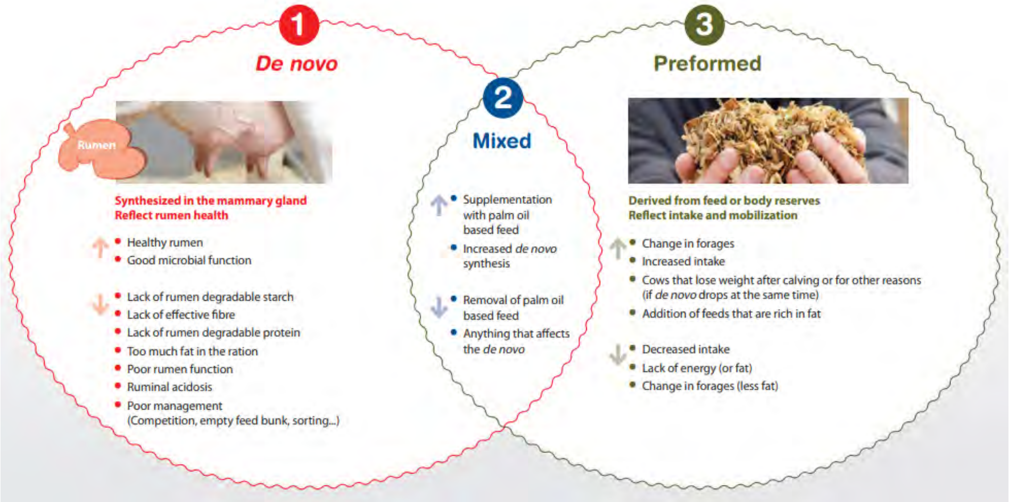


Figure 1. Milk FA groups based on origin and potential causes of increases or decreases of these groups in milk (g/100g milk).

Since April 2019, Lactanet offers routine milk FA profile analysis to all Quebec dairy producers, and more recently to some producers from other provinces in Canada as well. The objective is to implement this analysis across the labs in Canada and therefore be able to offer it across the country. Meanwhile, Quebec data was used to benchmark and compare FA results from herds based on different criteria, in order to provide a better understanding of this new metric.

Data analysis and main findings

Bulk tank milk samples from 3,200 dairy herds on milk recording in Quebec (N = 1.9M) analyzed from April 2019 to July 2022 were used in this analysis to evaluate the use of FA as a monitoring tool.

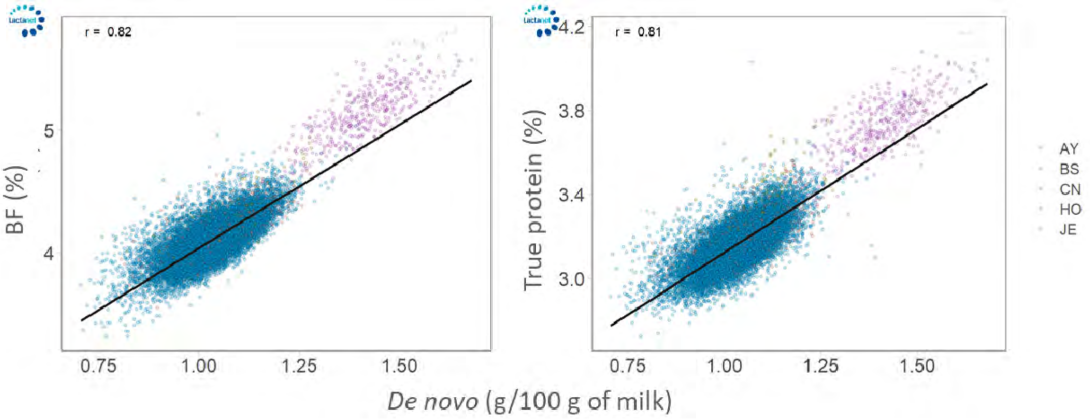


Figure 2. Relationship between de novo and main milk components (Butterfat. BF; and true protein) expressed in g/100g of milk for 1.9M bulk tank samples from 2019 to 2022 in Quebec dairy herds of different predominant breed type.

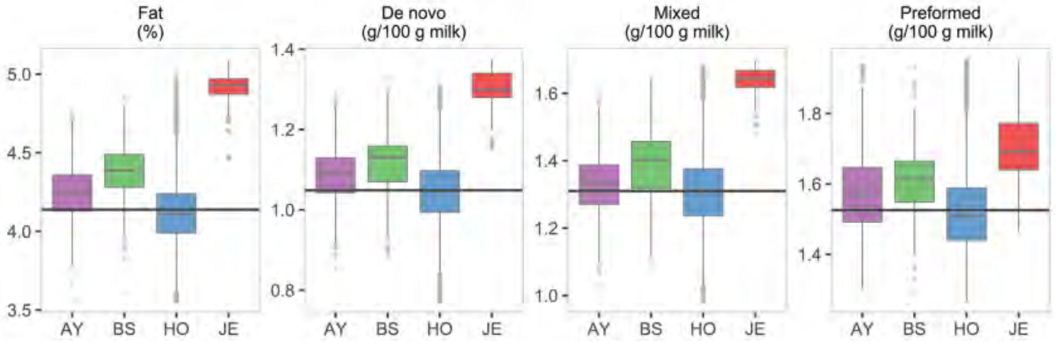


Figure 3. Boxplot representation of milk fat and fatty acid concentrations (g/100g milk) for 2821 Quebec herds of different breeds (AY = Ayrshire; BS = Brown Swiss; HO = Holstein; JE = Jersey). Horizontal line represents provincial average.

Milk fat and protein were correlated with de novo FA from bulk tank milk (rcorr=0.81–0.82; Figure 2) and to a lesser extent with preformed FA (rcorr=0.45 and 0.04, respectively). On a milk basis, each 0.1-percentage unit of increase in de novo FA increased milk fat content by 0.201 unit (r_{corr}=0.82) and milk protein content by 0.117 unit (r_{corr}=0.81); whereas each 0.1-unit increment in preformed FA increased milk fat by 0.099 unit (r_{corr}=0.45) and had no impact on milk protein (r_{corr}=0.04).

Fatty acid benchmarks

Several benchmarks were produced such as for milking system, organic status, herd size and milking frequency and are made available online (https://lactanet.ca/en/profilab-factsheets-with-benchmarks/). Main results are summarized below.

Bulk tank values were matched with the closest test-day results for milk recording herds (N=3700) in order to compare milk FA profiles of herds of different breeds or performance levels. On a milk basis (g FA/100g milk) Jersey (JE) herds had higher concentrations of all fatty acid groups, followed by other colored breeds and finally Holstein (HO) herds (Figure 3). Interestingly, the variation of FA concentrations within a breed was often higher than across different breeds, as shown on the previous Figure.

Holstein herds were classified in percentile ranking based on performance (kg milk/cow/day) to compare milk FA profiles of different production level groups (Figure 4). Again, results revealed bigger within-group difference than between-group variation. A similar comparison was performed with herd performance expressed as kg of milk fat/cow and day (Graph not shown). Overall, with increasing herd performance, *de novo* FA increased and preformed FA decreased. This trend was inversed for high performing herds (≥1.70 kg milk fat/cow and day), which further emphasizes the need to compare herds with specific milk FA profiles to their appropriate peers.

Organic status of herds and its seasonal impact on milk fat and fatty acids was also studied and revealed very distinct patterns for organic herds with a noticeable increased in preformed fatty acids and concurring decrease in de novo fatty acids coinciding with the onset of the pasture season (Figure 5).

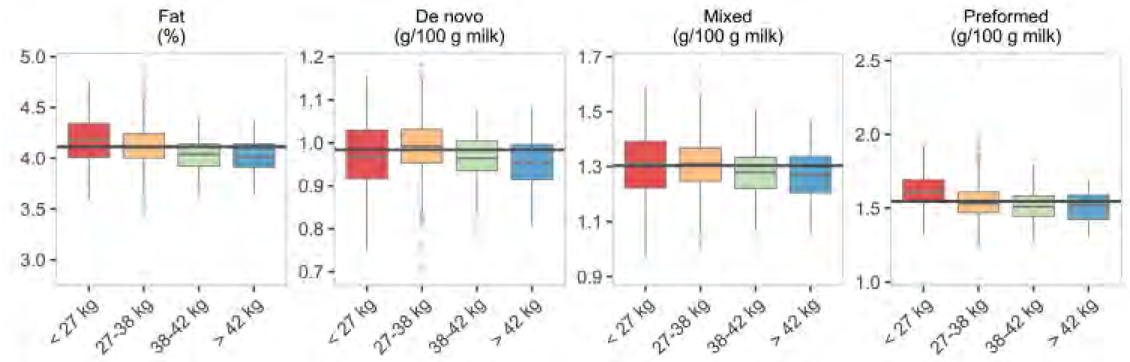


Figure 4. Boxplot representation of milk fat and fatty acid concentrations (g/100g milk) for 2652 conventional Holstein herds in Quebec, Canada based on their production level. Horizontal line represents provincial average.

A subset of 513 Holstein herds with available feed records was used to assess the impact of corn silage proportions in the ration on milk fat and fatty acid profiles. Herds were grouped based on the relative proportion of corn silage fed to the lactating herd (dry matter basis). Results suggest that increasing the proportion of corn silage in the ration up to 50% is associated with increased *de novo* and decreased preformed FA concentrations in milk (Figure 6).

Finally, the margin over feed costs (MOFC; in $ per kg of hl milk or per kg of milk fat) was greater with increased *de novo* FA and smaller with increased preformed FA, whereas no association was observed between the milk FA profile and MOFC per cow (results not shown).

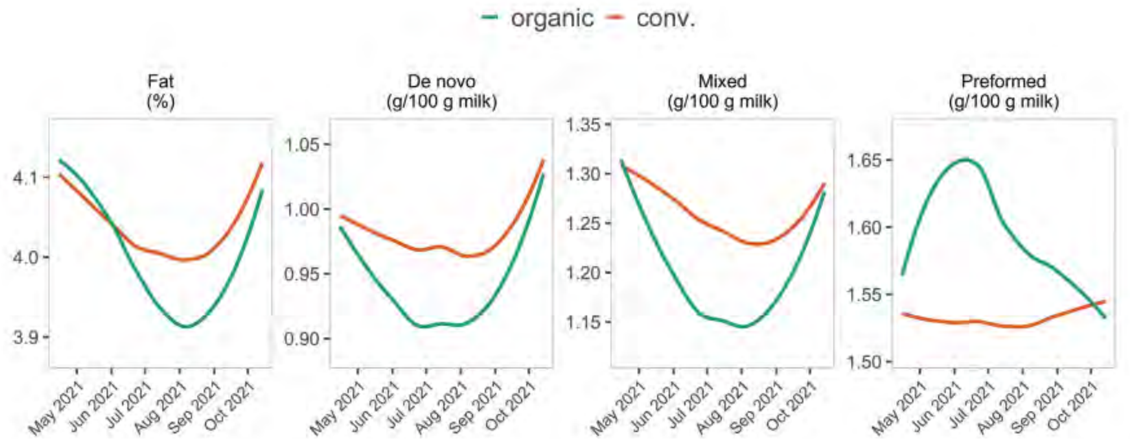


Figure 5. Seasonal variation of milk fat and fatty acid concentrations (g/100g milk) for 1572 conventional and 62 organic Holstein herds in Quebec, Canada.

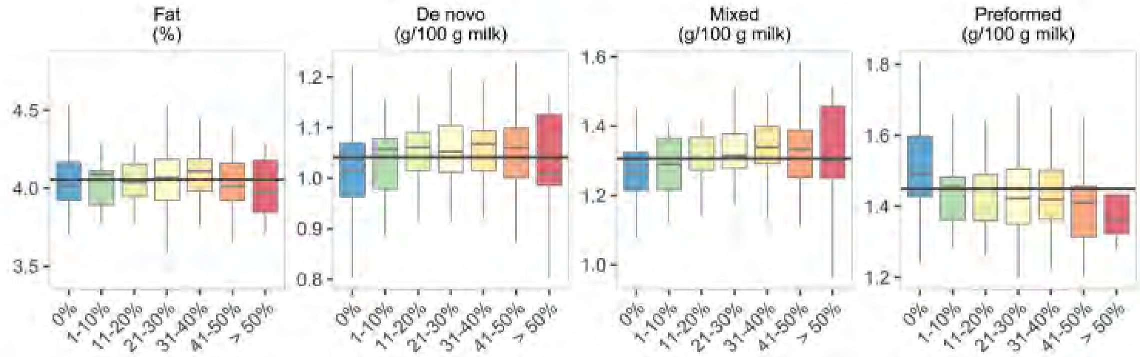


Figure 6. Boxplot representation of milk fat and fatty acid concentrations (g/100g milk) for 426 conventional Holstein herds in Quebec, Canada based on the proportion of corn silage in the ration (dry matter basis). Horizontal line represents provincial average.

Conclusion

In conclusion, variation of milk FA profiles within specific cohorts of herds is often greater than within cohorts, thus highlighting the importance for using herd-specific benchmarks in relation to historical farm data and strategic goals for a herd (internal benchmarking). An interactive dashboard application was developed to visualize milk FA profiles including benchmarks for the user to interpret their herd performance against their peers. Current usage in Quebec provides these values updated at every bulk tank pick-up, thus approximately every 48 hours. Future developments include the development of a cow-level monitoring tool as well as the use of artificial intelligence to provide support in the detection of variations and interpretation of milk fatty acid profiles at the herd level.

Acknowledgement

The authors wish to thank the Canadian and Quebec governments for financial funding through the Canadian Agricultural Partnership, Les Producteurs de lait du Québec, Dairy Farmers of Ontario, Lactanet lab colleagues in Ste-Anne-de-Bellevue, as well as Lactanet field advisors for their advice and time in discussing the results and interpretation on farm.

References

Woolpert, M.E., H.M. Dann, K.W. Cotanch, C. Melilli, L.E. Chase, R.J. Grant and D.M. Barbano. 2016. Management, nutrition, and lactation performance are related to bulk tank milk de novo fatty acid concentration on northeastern US dairy farms. Journal of Dairy Science 99: 8496-8497. http://dx.doi.org/10.3168/jds.2016-10998.

Woolpert, M.E., H.M. Dann, K.W. Cotanch, C. Melilli, L.E. Chase, R.J. Grant and D.M. Barbano. 2017. Management practices, physically effective fiber, and ether extract are related to bulk tank milk de novo fatty acid concentration on Holstein dairy farms. Journal of Dairy Science 100: 5097-5106. https://doi.org/10.3168/jds.2016-12046.

Optimize breeding and replacement decisions using milk recording data

R.H. Fourdraine[1], J.S. Clay[1] and A. De Vries[2]

[1]Dairy Records Management Systems, 313 Chapanoke Road,
Suite 100, Raleigh, NC 27603, USA
[2]University of Florida, 2250 Shealy Drive, Gainesville, FL 32611, USA

Abstract

With more breeding options available, decisions regarding which animals to replace and which animals to breed and the type of semen to use has become increasingly complex. US dairy farmers typically consider a couple key individual performance measures in the decision making process. Assessing the future economic returns would provide a better assessment, however requires more data inputs and a more complex method of analysis. This paper will explain how a prediction model developed by the University of Florida in collaboration with Dairy Records Management Systems will reduce the guesswork and simplify the process of making replacement and breeding decisions using cow performance data collected through milk recording.

Keywords: Fourdraine, Clay, De Vries, optimization, breeding, replacements, economic, decision, support, milk, recording.

Introduction

Sorted semen has enabled important changes in how producers manage programs for replacement animals while optimizing overall genetic progress. Increased use of sorted dairy semen can achieve more female calves to replace existing cows or to expand the herd without significantly impacting conception rates. A ready supply of replacement heifers provides more options to replace cows that are low producing, have reproductive problems or are not healthy. Genomic testing accelerates a herd's overall genetic progress when the highest genetic merit cows and heifers are bred with sorted dairy semen. Recently, in addition to conventional and sorted semen, U.S dairy farmers have added another option when making breeding decisions – semen from beef bulls. Typically, the most valuable animals (heifers and higher genetic merit first lactation cows) are bred to sorted dairy semen at first service. Conventional semen usually is used for later services. Because dairy farmers want to take advantage of higher revenues for a dairy/beef calf, there has been an increased use of beef semen in the dairy industry. Utilizing Dairy Records Management Systems (DRMS) breeding data from over 2 million cows, recent results found not only increased use of beef semen but there have been recent changes in the age and timing that dairy cows have been bred to beef semen.

With so many options available including using cows as embryo donors or recipients, how can a producer determine the best breeding option for each animal and thereby improve returns? Utilizing a prediction model developed by the University of Florida in collaboration with DRMS, decision support software was developed that reduces the guesswork while simplifying the process of making replacement and breeding decisions using cow performance data collected through milk recording.

This new decision support software features individual cow genetic information coupled with cow performance data to predict each cow's future performance and economic values for replacement and mating decisions. This information can be used to decide whether a cow should be retained, bred or replaced. In addition, the model assigns an economic value to each of the three potential mating sire choices: sorted dairy semen, conventional dairy semen or beef semen. As inputs related to each cow's stage of lactation, reproductive status or genetic information may change daily, outcomes from the model will change as well, and, replacement and breeding decisions can be made based on the most current data.

U.S. breeding trends

Using data from over 10 million DRMS breeding records, Figure 1 illustrates the type of semen used for Holstein and Jersey breedings during the recent two years. Based on this two-year period, it is clear that there has been an increase in use of both beef semen and dairy sorted semen that has resulted in reduced use of conventional dairy semen. Overall, almost 50% of breedings to Holstein and Jersey cows were to sorted dairy or beef semen.

With easier access to replacement animals and improved fertility, producers have lowered the threshold in terms of how many times they will breed a cow before deciding to stop breeding and to replace her. Therefore, as the number of times a producer will

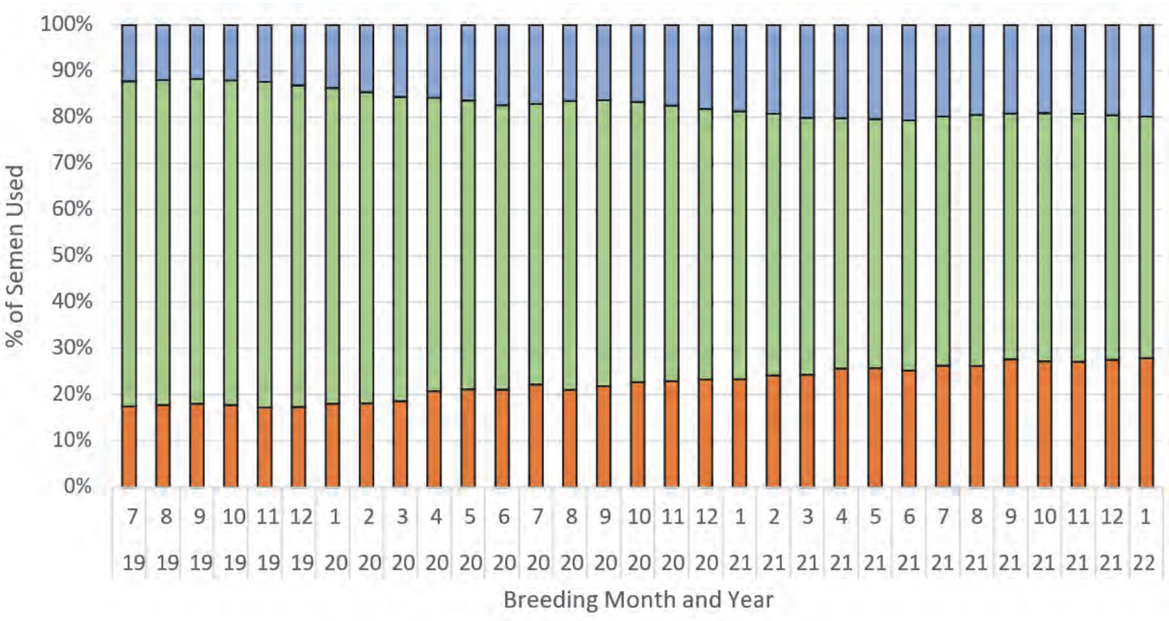


Figure 1. Semen usage trend for Holstein and Jersey cows.

Table 1. Semen distribution for Holstein cows by parity, service number, and type of semen.

	Service #	% Beef	% HO Conv	% HO Sorted	% Jersey
Heifer	1	4.0%	36.0%	58.3%	1.7%
	2	6.7%	40.0%	51.7%	1.6%
	3	25.3%	48.2%	24.6%	1.9%
1st Lact	1	13.1%	59.8%	24.8%	2.3%
	2	19.5%	61.2%	17.0%	2.2%
	3	37.7%	55.3%	6.3%	0.7%
2nd Lact	1	24.8%	60.6%	13.4%	1.1%
	2	29.8%	60.4%	9.0%	0.9%
	3	42.8%	53.7%	2.8%	0.6%
3> Lact	1	36.1%	55.2%	7.8%	0.8%
	2	39.8%	54.4%	5.2%	0.6%
	3	50.2%	47.1%	2.1%	0.7%

breed a cow before considering culling has dropped in the past five years, the decision process to breed with three choices of semen at hand has changed significantly. A fourth option for herds that consider embryo transfers is to use a low genetic merit cow as a recipient. Table 1 shows a comparison based on parity and service number and the type of semen used on Holstein Cows in the DRMS breeding records

Over fifty percent of US Holstein heifers were bred to sorted semen on first or second service. This makes sense because these should be genetically superior animals and because breeding these animals with sorted semen should yield the greatest number of female offspring to use as replacements. Interestingly by the third service, the percent of breedings to beef semen on heifers is twenty-five percent. Additionally, the use of beef semen increases rapidly with lactation number. Thirteen percent of first lactation cows were bred to beef semen on first service. Use of beef semen increases to thirty-six percent for third and greater lactation cows.

Another observed difference relates to semen usage based on herd size. Figure 2 show the distribution of semen usage for herds with a minimum of 1,000 cows. Based on the DRMS data, smaller herds used a substantively lower percentage of beef semen and sorted semen than larger herds. This trend for smaller herds did not change much in the recent two years while the trend in larger herds was to use more beef and sorted dairy semen and less conventional dairy semen.

Over fifty percent of US Holstein heifers were bred to sorted semen on first or second service. This makes sense because these should be genetically superior animals and because breeding these animals with sorted semen should yield the greatest number of female offspring to use as replacements. Interestingly by the third service, the percent of breedings to beef semen on heifers is twenty-five percent. Additionally, the use of beef semen increases rapidly with lactation number. Thirteen percent of first lactation cows were bred to beef semen on first service. Use of beef semen increases to thirty-six percent for third and greater lactation cows.

Another observed difference relates to semen usage based on herd size. Figure 2 show the distribution of semen usage for herds with a minimum of 1,000 cows. Based on the DRMS data, smaller herds used a substantively lower percentage of beef semen and sorted semen than larger herds. This trend for smaller herds did not change much in the recent two years while the trend in larger herds was to use more beef and sorted dairy semen and less conventional dairy semen.

Dairy managers face an increasingly complex decision making process pertaining to when a cow should be sold, bred and which sire to breed the cow to. In most cases, decisions are based on a small number of variables that can be easily obtained on an

Need for a breeding and culling decision support tool

individual cow. For example, a replacement decision typically is based on how much did the cow produce in her current or past lactation, did she experience several health problems, is she open and how many times was she bred? The breeding decision concerning which semen type to for a cow will include some of the same variables but in most cases also includes the genetic value of the animal. Although this process can be managed using a report from herd management software and making a personal judgement, it does not take into account some of the more complex factors such as milk price, replacement cost, value of a dairy female versus dairy male or the value of a dairy beef cross calf, to name a few.

To make more informed decisions and allow for a larger number of parameters for individual cows and the herd to be part of the analysis, the University of Florida Animal Sciences department, in collaboration with DRMS, developed a model that will calculate an economic value of each the cow in the herd and also calculates values for breeding a cow to either dairy conventional, dairy sorted or beef semen. The outcomes from the model were then incorporated into a decision support tool showing the relative ranking of the cow in the herd and provides a first choice and second choice breeding recommendation for cows that need to be bred.

Description of the model

Data inputs to the model consist of parameters for the herd and individual cows. Most of the herd parameters such as conception rates, milk production levels by lactation can be obtained from the on farm PCDART herd management software or monthly DRMS test day processing. Financial information must be provided by the producer. Financial information consists of milk price (including milk components), heifer replacement costs, beef price, selling price of a dairy bull calf and dairy x beef calf, and costs per unit of sorted, conventional dairy and beef semen. The model also allows for a feed cost adjustment.

Cow data includes cow lactation number, days in milk, status data (milking or dry), reproductive status (Bred, Pregnant, Open, etc..), the last 2 test day milk, fat, and protein weights, current lactation to date milk, fat and protein weights, and past lactation 305D milk, fat and protein weights. The economic genetic merit (Net Merit Dollars) is also included for each cow.

The model estimates future cash flows following each decision alternative (keep, replace, and breed with sexed dairy, conventional dairy, or beef semen) following optimal replacement and breeding decisions now and into the future. These future cash flows are calculated using dynamic programming and future cow performance estimates based on the milk recording data and prices. Cash flow estimates include cash flows from replacement heifers when the cow is replaced now or in the future. Typically cash flows for six years into the future are needed to capture all changes in cash flows that follow from a replacement or breeding decision. The model automatically determines this length of time.

A novel feature of this model is that future cash flows are calculated for each cow based on her best estimated future performance for milk production, fertility, forced culling, genetic merit etc. This allows for accurate future cash flow estimates. Finally, the model calculates the differences between the net present values of the future cash flow estimates and presents them as four different economic values. For example, the economic value of the cow (Keep Dollar) is the difference between the net present value of the keep (and optimal breeding) decision and the decision to replace the cow now with a heifer. When this economic value is greater for a cow, the more valuable the cow is to keep in the herd. Cows that rank low for this economic keep value should be culled and replaced. Insemination values are calculated as the net present value of breeding with that type of semen compared to the net present value of the decision to delay breeding until the next opportunity.

Presentation of results

As stated above, the model returns four key economic values, the economic value of the cow (Keep Dollar), plus three economic insemination values for breeding the cow to either Sorted Dairy, Conventional Dairy or Beef semen respectively. The results from the model are merged with other cow data, and the percentile ranking for the cow's Keep Dollar that provides an easily interpreted rank of relative economic worth for each cow within the herd.

The three insemination values are ranked for each cow and the recommendation with the highest economic merit is presented as first choice, while the next highest value is presented as second choice. The dollar difference between first and second choice is provided to determine the economic difference between first and second choice. Picture 1 provides an example output for cows that are due to be bred.

In the example above, Cow 8068 has a Keep Dollar of 1020 which ranks her at 66th percentile (Keep Perc.) within the herd. The first choice recommendation is to breed her with (S)orted dairy semen and second choice would be (C)onventional dairy semen. Breeding 8068 with Sorted semen has a 17 dollar advantage over breeding her with conventional dairy semen. Cow 8075 ranks a much lower 37th percentile so the breeding recommendation is to breed her with (B)eef semen. However, her second choice of (C)onventional semen is close behind with a 1 Dollar difference.

Dairy farmers can use the combination of the cow's Keep percentile and first semen choice recommendation to determine which cows to breed (low Keep percentile cows should be considered for replacement). If semen inventory is limited, ranking cows by first choice and the economic difference between first and second will allow the dairy farmer to select the highest dollar difference values that would be the cows with the strongest recommendation to breed to first choice.

Conclusion

This new decision support tool will help dairy farmers make more informed breeding and culling decisions. The model is customizable using each herd's unique situation, and, new parameters that would impact a cow's economic value can be added. In addition, this new tool is time sensitive and will use the most current data available from the herd management software to ensure changes in production or reproductive status are reflected in a cow's relative economic value in the herd. Current efforts are focused on testing the model with several dairies, and, future enhancements will include breeding

Table 2. Sample Output.

ID	Keep Dollar	Keep Perc.	First choice	Second choice	First over Second
8066	729	43	C	S	2
8068	1020	66	S	C	17
8071	739	44	B	C	4
8072	863	54	S	C	18
8074	875	56	S	C	10
8075	646	37	B	C	1

recommendations for heifers allowing further expansion to a herd level decision tool taking into account number of replacements needed to maintain herd size.

Milk fatty acid profiles in early lactation as potential indicators for reproduction success

D. Warner[1], L. Fadul-Pacheco[1], R. Gervais[2] and D.E. Santschi[1]

[1]Lactanet, 555, boul. des Anciens-Combattants, Sainte-Anne-de-Bellevue, H9X 3R4, Québec, Canada
[2]Département des sciences animales, Université Laval, 2425, rue de l'Agriculture, Québec, G1V 0A6, Québec, Canada
Corresponding Author: dwarner@lactanet.ca

Abstract

The reproductive performance of a dairy cow can be influenced by her metabolic status during early lactation. As this period is also characterized by important changes in the milk yield and milk fatty acid (FA) profile, we aimed to assess the potential of first test date milk FA profiles analysed by Fourier-transform infrared (FTIR) spectroscopy to predict the subsequent reproduction performance. First-test date records from 246,345 Holstein cows in early lactation (5–35 DIM) across Quebec, Canada, were included in the analysis. Records pertained to 2,835 herds and spanned over 2 years (2020 and 2022). Cows were clustered in similar cohorts based on their first test date milk composition, FA (expressed as g/100 g of total FA) and yield using principal components. Clustering was based on the CLARA concept with a k-medoid approach on subsamples to overcome the limitation of computational resources in clustering a large dataset. Three clusters were identified after iteration. Reproduction success was assessed based on the interval between first service to conception (FSTC) and culled by 60 DIM (CULL) using a linear regression and binomial logistic regression mixed effect model, respectively. Cluster 3 was composed of cows with high preformed FA (56.2% on total FA basis), 18:1-to-14:0 ratio (4.74), BHB (0.17 mM) and fat-to-protein ratio (1.51), but low de novo FA (17.1%). In contrast, Cluster 2 was composed of cows with low preformed FA (39.9%), 18:1-to-14:0 ratio (2.08), BHB (0.08 mM) and fat-to-protein ratio (1.17), but high de novo FA (25.8% total FA). Cluster 1 was in between Cluster 3 and 2 and had the highest milk yield. The FSTC was greatest ($P < 0.001$) for Cluster 3 (62.2 days) followed by Cluster 1 (60.7 days) and Cluster 2 (57.9 days). Likelihood for CULL was greatest ($P < 0.001$) for Cluster 3 (odds ratio of 2.6). This preliminary analysis suggests that FTIR milk FA profiles at first test date could be used as early indicators for the following reproduction success of dairy cows and help improve the transition management through continuous monitoring.

Keywords: Decision support, DHI, fertility, early lactation.

Introduction

The increased mobilization of body reserves after parturition is reflected by changes in milk composition and milk fatty acid (FA) profiles, in particular through increased uptake of FA by the mammary gland and decreased de novo synthesis of FA. As early lactation is also a critical phase for later production and reproduction performance with metabolically challenged cows more likely to suffer from early lactation disorders, first test date milk composition might be used as early indicator for reproductive success. The objective of this study was to assess whether first test milk components and FA profiles are associated with the subsequent reproductive performance in commercial dairy herds using a large DHI dataset.

Material and methods

Test day records with milk components, milk and reproduction performance were obtained by Lactanet database (Lactanet, Sainte-Anne-de-Bellevue, Canada). For this retrospective study, we restricted the dataset to first test date records from Holstein cows in early lactation (5-35 DIM) with test results over two complete years (2020 and 2021). The dataset included 246,345 cows (30% in parity 1; 26% in parity 2; and 44% in parity 3 or more). In total, 2,835 dairy herds from Quebec, Canada, equipped with a pipeline (N = 2,226), parlour (N = 251) or automatic (N = 358) milking system were included in the study. As a confirmed gestation was not recorded, the subsequent lactation start date was used to define the actual conception date. Therefore, for reproduction performance indicators requiring a confirmed gestation, the dataset was limited to 28,254 test dates from 100,711 cows with tests dates until July 2021.

Milk samples were analysed as regular DHI milk samples for fat, protein, lactose, MUN, SCC, and FA by Fourier-transform infrared (FTIR) using MilkoScan FT+ and MilkoScan 7 RM instruments (Foss, Hillerød, Denmark). Fatty acids included individual and groups of FA (C14:0, C16:0, C18:0, and sum of C18:1, saturated, mono-unsaturated, poly-unsaturated, short-chain, medium-chain, long-chain, de novo, mixed, and preformed FA) as described in FOSS Application Note 64 and Schwarz *et al*. (2018).

Data were cleaned to remove implausible values and outliers by removing data below the 1st and above the 99th percentile. The milk FA profile was expressed as % of total FA (TFA; i.e., the sum of de novo, mixed, and preformed FA).

Cows were grouped in similar cohorts based on their first test date milk composition and yield using k-medoid clustering via principal components. The principal components were used to de-noise the data and to balance the influence of similar milk components. To deal with the large number of observations a sampling approach was used for clustering (CLARA; Kaufman and Rousseeuw, 1990). The procedure consists of randomly splitting the dataset into multiple subsamples and applying the PAM algorithm (Partitioning Around Medoids; Kaufman and Rousseeuw, 1990) to generate the optimal set of cluster centres for each subsample, here computed based on a dissimilarity function based on the Euclidean distance. The optimal number of clusters was evaluated using the average Silhouette approach.

Differences in milk composition and reproduction performance among the identified clusters were evaluated through a mixed effect linear regression with fixed effects assigned to cluster, milking system, lactation, DIM at test date, season and year, and random effects assigned to herd. For a binary outcome, a mixed effect logistic regression was used. All analyses were conducted using R (version 4.1.3; R Foundation for Statistical Computing, Vienna, Austria) and add-on packages computing the principal components (FactoMiner; Lê *et al*., 2018), clusters (cluster; Mächler *et al*., 2022), and mixed effect linear regression (lmerTest; Kuznetsova *et al*., 2017), and mixed effect logistic regression (lme4; Bates *et al*.; 2015).

Results and discussion

Clustering animals based on their first test milk composition, yield, and fatty acid profiles resulted in three distinct clusters. Most cows were assigned to Cluster 1 but herds, herd size and housing system were similarly distributed among clusters (Table 1).

The milk composition differed among clusters ($P < 0.001$; Figure 1) which resulted in Cluster 3 having the least ideal profile with a high fat-to-protein ratio (FPR; mean of 1.51), high BHB (0.17 mmol/L), high preformed FA (56.2% of TFA), a high C18:1-to-C14:0 ratio (4.74), low de novo FA (17.1% of TFA), and low mixed FA (26.8% of TFA). Although the FPR is typically high for early lactation cows due to the higher milk fat synthesis during the postpartum negative energy balance, cows with a FPR value greater than 1.5 (Heuer *et al*., 1999) or 2.0 (Toni *et al*., 2011) may show an increase in postpartum diseases. Likewise, cow with milk BHB above 0.20 mmol/L were more likely to suffer from hyperketonemia (Denis-Robichaud *et al*., 2014), and cows with de novo FA below 20% of TFA were more likely to have early-lactation diseases and be removed from the herd (Bach *et al*., 1999). Milk composition for Cluster 2 was particular high in de novo (25.8% of TFA) and mixed (34.5% of TFA) FA, low in preformed FA (39.7% of TFA) and C18:1 to C14:0 (2.08). Cluster 1 was intermediate among the three clusters but had a higher milk yield (+9.1 kg milk to Cluster 2).

Table 1. Cluster description.

	Cluster		
Variable	**1**	**2**	**3**
Number of herds	2722	2726	2717
Number of cows	107,657	72,889	57,913
Number of lactating cows (median)	71.7	71.9	67.9
Number of herds in tie stalls	2,195	2,200	2,194

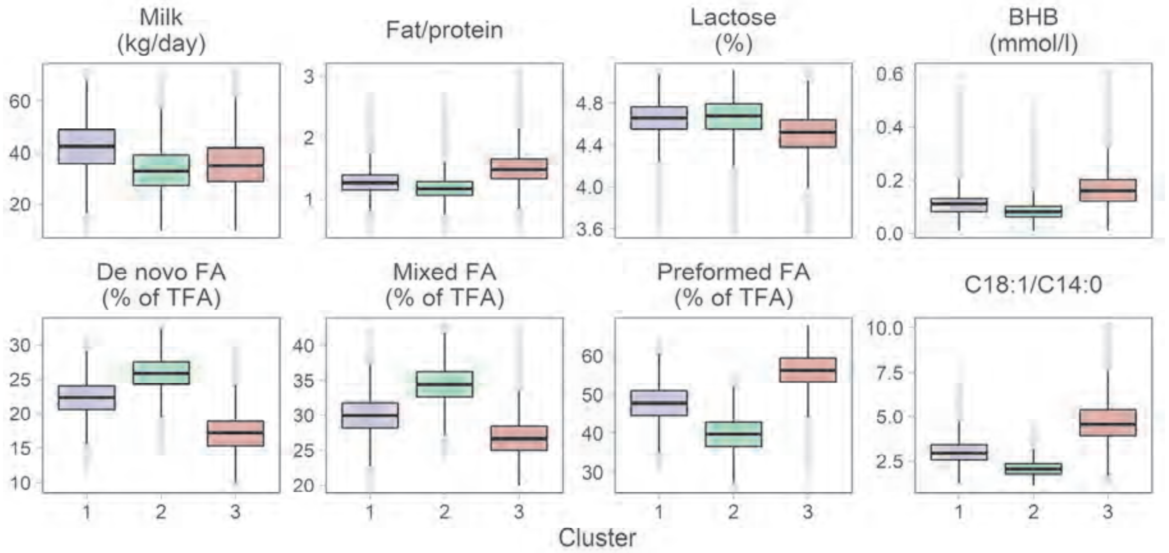


Figure 2. Milk composition and milk fatty acid (FA) profile at first test date by cluster.

Preliminary analyses suggested that the reproduction performance differed among clusters. Cluster 3 with the least ideal milk composition and FA profile was associated with the longest ($P < 0.001$) interval between first breeding to conception at 62.2 days (confidence interval of 32.2-92.3 days) compared to 57.9 (27-8-87.9) days for Cluster 2 and 60.7 (30.6-90.8) days for Cluster 1. Likewise, cows assigned to Cluster 3 were more likely ($P < 0.001$) to be culled before 60 DIM than cows in Cluster 2 (odd ratio of 2.6 and 1.3, respectively). The contrast among clusters is expected to be even more greater considering that cows left before the event occurred were not considered for first breeding to conception (e.g., cows removed before the first breeding or before the start of the next lactation). Future studies should therefore use a survival analysis approach to take into account censored observations in the analyses.

These findings suggest that the milk composition and milk FA profile at first test date can be indicative for the following reproduction success. As such, these findings can be useful for prevention measures and can help with transition management and decision-making support for future lactations but will be of limited use for current early lactation cows due to the monthly DHI test sampling scheme generally used in practice. A more frequent test sampling during the first two weeks of lactation should be considered for a timely intervention. Likewise, test day information from the previous lactation might be relevant for the current reproduction performance and add predictive power to forecast future reproduction issues.

Acknowledgement

This project was funded by the Natural Sciences and Engineering Research Council of Canada (Ottawa, ON, Canada) and the Consortium for Research and Innovation in Industrial Bioprocesses in Quebec (CRIBIQ; Québec, QC, Canada).

References

Bach, K.D., D.M. Barbano and J.A.A. McArt. 2019. Association of mid-infrared-predicted milk and blood constituents with early-lactation disease, removal, and production outcomes in Holstein cows. Journal of Dairy Science 102: 10129-10139. https://doi.org/10.3168/jds.2019-16926.

Bates, D., M. Mächler, B. Bolker and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67(1): 1-48. doi: 10.18637/jss.v067.i01.

Denis-Robichaud, J., J. Dubuc, D. Lefebvre and L. DesCôteaux. 2014. Accuracy of milk ketone bodies from flow-injection analysis for the diagnosis of hyperketonemia in dairy cows. Journal of Dairy Science 97: 3364-3370. http://dx.doi.org/ 10.3168/jds.2013-6744.

Heuer, C., Y.H. Schukken and P. Dobbelaar. 1999. Postpartum body condition score and results from the first test day milk as predictors of disease, fertility, yield, and culling in commercial dairy herds. Journal of Dairy Science 82: 295-304. https://doi.org/10.3168/jds.S0022-0302(99)75236-7.

Kaufman, L. and P. Rousseeuw. 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley and Sons, Inc., New York. doi: 10.1002/9780470316801.

Kuznetsova, A., P. B. Brockhoff and R.H.B. Christensen. 2017. lmerTest Package: Tests in linear mixed effects models. Journal of Statistical Software 82(13): 1–26. doi: 10.18637/jss.v082.i13.

Lê, S., J. Josse and F. Husson. 2018. FactoMineR: An R package for multivariate analysis. Journal of Statistical Software 25: 10.18637/jss.v025.i01.

Mächler, M., P. Rousseeuw, A. Struyf, M. Hubert and K. Hornik. 2022. cluster: Cluster Analysis Basics and Extensions. https://CRAN.R-project.org/package=cluster.

Schwarz, D., M. Rosenberg Bak and P. Waaben Hansen. 2018. The New FOSS Fatty Acid Origin Package – Basics Behind the Prediction Models. FOSS White Paper.

Toni, F., L. Vincenti, L. Grigoletto, A. Ricci and Y.H. Schukken. 2011. Early lactation ratio of fat and protein percentage in milk is associated with health, milk production, and survival. Journal of Dairy Science 94: 1771783. https://doi.org/10.3168/jds.2010-3389.

Using box time indicators to rank cows according to their efficiency in robotic milking systems

L. Fadul[1], R. Cue[2], G. Bisson[1], R. Moore[1] and R. Lacroix[1]

[1]Lactanet, 555 Boulevard des Anciens-Combattants, H9X 3R4,
Sainte-Anne-de-Bellevue, Quebec, Canada
Corresponding Author: lfadul@lactanet.ca
[2]McGill University, 21111 Lakeshore Road Sainte-Anne-de-Bellevue,
H9X 3V9, Quebec, Canada

Abstract

Large amounts of data are generated by robotic milking systems (RMS), creating an opportunity to add-value to these data by developing tools that can help improve decision making and RMS management. One of the strategies to improve RMS efficiency is to select cows based on their performance at the RMS (kg of milk produced per total time spent in the RMS; kg milk/minute of box time). The objective was to develop box time-based indicators to rank the cows according to their performance at the RMS. Data was collected for 8 months (October 5, 2020, until June 14, 2021) from 41 RMS Holstein herds, for a total of 5,429 cows and 411,405 records. Data included milk production, box time, preparation time, DIM, parity, milking start and end time, milk composition and milk prices. Calculations were done based on the ICAR standards (i.e., 4-day average). Preliminary analysis has shown that box time and hence the indicators vary significantly across and within lactation. Variability was very high during early lactation, presumably associated with the important increase of production during that period. The variability was also very high in late lactation, because of a decreasing number of observations due to ended lactations, thus data were limited to DIM between 28 and 365. Regression analysis was performed to generate factors to correct indicators for parity groups (1,2 and +3) and stage of lactation (DIM). Herd was considered as a random effect. The best fittings were obtained with quadratic polynomials. The resulting equations can be used to generate a factor to correct the data to 150 DIM within each parity, and to adjust to parity 2 for data in the other two parity groups. In addition to the amount of milk produced per minute of box time an economic indicator was also developed: dollars per minute of box time. The average of box time was 7.22 min/cow, and the average of the indicators were 1.86 kg milk/min of box time and 1.50 Can$/min of box time. The box time indicators can be used to rank and subsequently select cows according to their performance at the RMS, in addition to other indicators such as somatic cell count. They can also be used to calculate benchmarks for comparative analysis across RMS herds, as the RMS efficiency is key for the economic success of the farms.

Keywords: Box time, robotic milking systems, efficiency.

Introduction

Data produced by robotic milking systems (RMS) is an opportunity to personalise the milking management, as the RMS generates a large amount of individual cows' data. These data also allow the development of tools that can help the overall RMS management. One of the key factors of RMS management and hence the successful economic use of the robots is to improve their efficiency, (i.e., maximise the amount of milk produced per robot per day) (Carlstroem *et al*., 2013). Studies had suggested that the most important factor to achieve this is the total of kg of milk produced per total time spent in the RMS; kg milk/minute of box time (Carlstroem *et al*., 2013; Vosman *et al*., 2014). Box time is composed of the milking time and the preparation or treatment time (e.g., cleaning, attachment etc.), in other words is the time the cow spends in the robot and is expressed per milking or can be aggregated by day.

Genetics is also an important factor of RMS efficiency as milking time is related to cows' genetics (e.g., milkability) and preparation time to a lesser extent too (e.g., udder conformation and cows' temperament). However, it is known that genetic decisions have a medium to long term impact. For short term decision-making using data from the RMS is an opportunity. Nevertheless, in the future the integration of genetic data and RMS data could help improve the decision management tools. The objective was to develop box time-based indicators to rank the cows according to their performance at the RMS.

Material and methods

Data was collected for 8 months (October 5, 2020, until June 14, 2021) from 41 RMS Holstein herds, for a total of 5,429 cows and 411,405 records. Data included milk production, box time, preparation time, DIM, parity, milking start and end time, milk composition from the RMS sensors and milk prices. Calculations were done based on the ICAR standards (i.e., 96 hours average). The RMS efficiency per cow was calculated as the average of milk production (kg) per minutes of box time for a 96h period. To have an economic box time indicator the dollar per minute of box time was calculated as follows: dollar value of milk produced per minute of box time for a 96h period; where the dollar value of milk was calculated as a linear combination of milk component yields. Milk components were available from the RMS sensors and a constant was used for lactose content.

Our preliminary analysis has shown that box time and hence the indicators vary significantly across and within lactation (Figure 1a), which has been also reported by Heringstad and Bugten (2014). The variability was also very high in late lactation, because of a decreasing number of observations due to ended lactations, thus data were limited to DIM between 28 and 365. Regression analysis was performed to generate factors to correct indicators for parity groups (1,2 and +3) and stage of lactation (DIM). Herd was considered as a random effect. Analyses were conducted in SAS version 9.4 (SAS Institute Inc.).

Results and discussion

The best fittings were obtained with quadratic polynomial equations. The resulting equations can be used to generate a factor to correct the data to 150 DIM within each parity, and to adjust to parity 2 for data in the other two parity groups (Figure 1a, b). The average of box time was 7.22 min/cow, and the average of the indicators were 1.86 kg milk/min of box time and 1.50 Can$/min of box time.

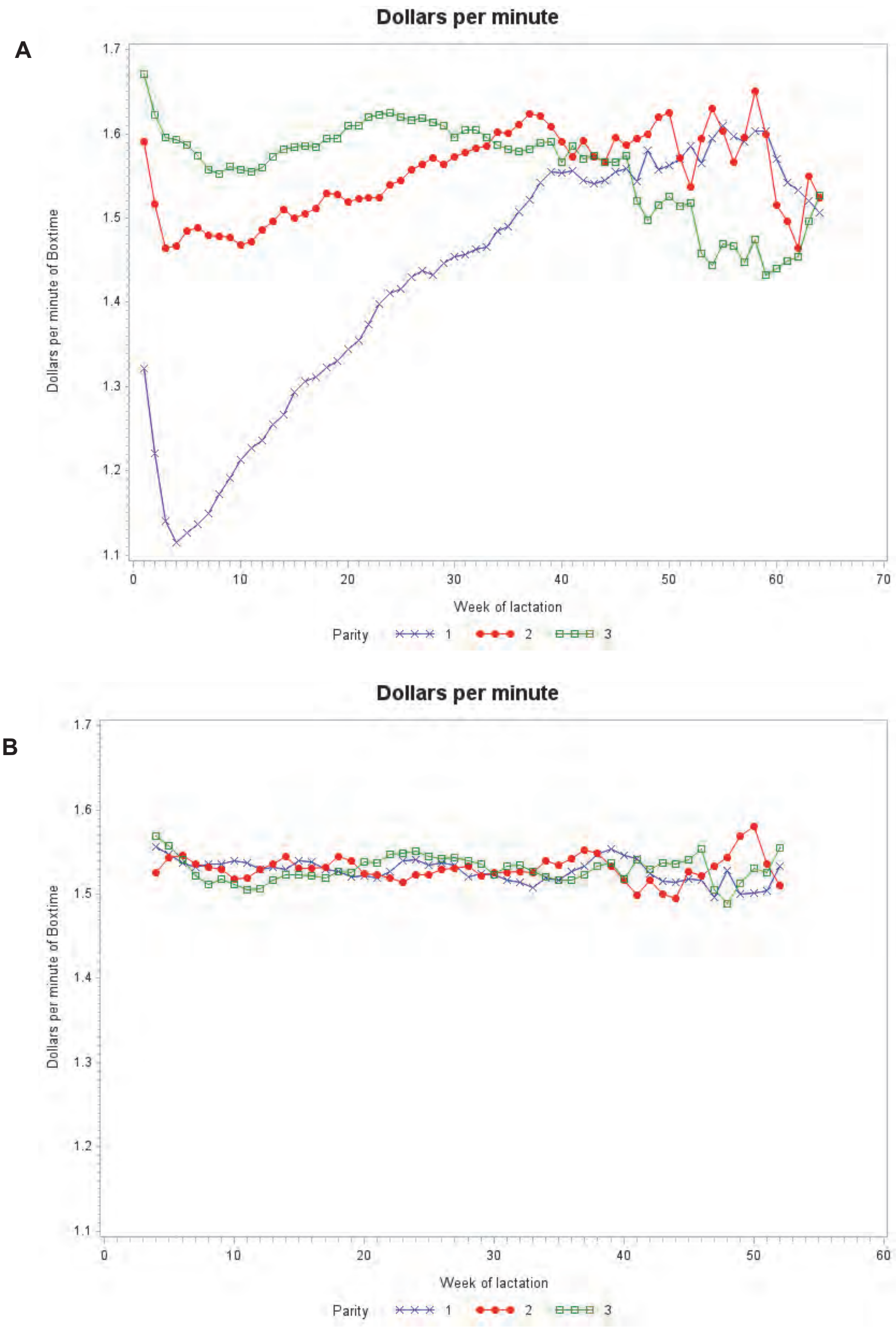


Figure 1. a) Raw data of the indicator dollar per minute of box time by parity groups (1,2, +3). b) Adjusted values of dollar per minute of box time by parity groups (1,2, +3). The DIM were converted to weeks in milk to facilitate the visualisation of the different points.

Studies have reported the importance of RMS efficiency in the management of robotic milking (Carlström *et al*., 2013; Heringstad and Bugten, 2014). Therefore, developing an indicator to rank the cows according to their performance has become a valuable decision support tool for RMS management, as shown by the implementation of these indicators among others in the interactive robot report by Lactanet.

Selecting cows for improving efficiency using the milkability trait could be a complementary approach. Milkability heritability ranges between 0.21 and 0.44 (Carlstroem *et al*., 2013; Cesarani *et al*., 2021) depending on the breed.

Conclusion

The box time indicators can be used to rank and subsequently select cows according to their performance at the RMS, in addition to other indicators such as somatic cell count. They can also be used to calculate benchmarks for comparative analysis across RMS herds, as the RMS efficiency is key for the economic success of the farms. Future work will focus on the possibility of adding genetic tools to refine the box time indicators and on the development of the adjustment factors for different breeds, as sufficient data becomes available to calculate robust adjustment factors.

Acknowledgement

Ce projet est financé en vertu de l'Accord Canada-Québec de mise en œuvre du Partenariat canadien pour l'agriculture. Ensemble, le gouvernement fédéral et le gouvernement du Québec ont investi 293 millions de dollars répartis sur une période de 5 ans, soit de 2018 à 2023.

Cet accord appuie des initiatives stratégiques qui aideront les secteurs à croître, à innover et à prospérer.

References

Carlstroem C, G. Pettersson, K. Johansson, E. Strandberg, H. Stålhammar and J. Philipsson, 2013. Feasibility of using automatic milking system data from commercial herds for genetic analysis of milkability. Journal of Dairy Science 96, 5324–5332.

Carlstroem C, E. Strandberg, K. Johansson, G. Pettersson, H. Stålhammar and J. Philipsson, 2016. Genetic associations of in-line recorded milkability traits and udder conformation with udder health. Acta Agriculturae Scandinavica, Section A — Animal Science 66, 84–91.

Cesarani A, A. Garcia, J. Hidalgo, L. Degano, D. Vicario, N.P.P. Macciotta and D. Lourenco, 2021. Genomic information allows for more accurate breeding values for milkability in dual-purpose Italian Simmental cattle. Journal of Dairy Science 104, 5719–5727.

Gäde S, E. Stamer, J. Bennewitz, W. Junge and E. Kalm, 2007. Genetic parameters for serial, automatically recorded milkability and its relationship to udder health in dairy cattle. Animal 1, 787–796.

Heringstad B and H.K. Bugten, 2014. Genetic evaluations based on data from automatic milking systems. In Proceedings, 10 th World Congress of Genetics Applied to Livestock Production (WCGALP). Vancouver, Canada.

Vosman J, G. de Jong and H. Eding, 2014. Breeding of Cows Suitable for an Automatic Milking System. In Interbull Bulletin. Berlin, Germany.

Automated anomaly detection for milk components and diagnostics in dairy herds

M. Ayat, G. Bisson, J. Prince, V. Fuentes, D. Warner, D.M. Lefebvre, D.E. Santschi, and R. Lacroix

Lactanet, 555 Boul des Anciens-Combattants, Sainte-Anne-de-Bellevue, Canada
Corresponding Author: rlacroix@lactanet.ca

Abstract

Changes in the bulk tank milk component profiles (e.g., basic components, MUN and specific fatty acids) often indicate some positive or negative effects of nutrition, management, and environmental factors. Generally, producers and their advisors detect abnormal trends by visual inspection of components on a report, which is tedious and requires skills due to large volume of information. Our interim goal is to develop analytical tools to assist in the identification of unwanted, abnormal trends. Our broad goal is to develop tools that can propose plausible diagnostics and possible actions to re-establish the desired situation.

To achieve the interim goal, identification of anomalies is based on three numerical indicators (deviation, variation, and gradient) and breed-specific population ranking. Deviation is the difference between component values and population benchmarks, while variation and gradient indicate short-term (e.g., 4 days) and longer-term (e.g., 10 days) changes, respectively. We developed a python package which is executed daily to generate indicator values for all milk components, calculate the herd rank for each indicator, and create a series of reports for validation purpose. Among all possibilities that were investigated to trigger alerts and attention messages, an approach based on extreme gradients of the three main groups of fatty acids was adopted to start with.

With respect to the broad goal, we explore a rule-based expert system approach for diagnosing and recognizing potential issues regarding herd or rumen health management. Currently, we are at the stage of eliciting, implementing, and validating diagnostic IF-THEN rules with the collaboration of domain experts and computer scientists. The rules are based on the anomaly indicators and ranks. In parallel, we are exploring the use of ontology and symbolic artificial intelligence to develop a more complete diagnostic and recommendation system.

Keywords: Bulk tank milk component, anomaly detection, time series analysis, rule-based system, diagnostic rule, ontology.

Introduction

Changes in the value of indicators such as bulk tank milk components generally indicate some positive or negative effects of management and environmental factors such as feed quality, feeding behaviour, or ambient conditions. The data provided on milk profiles keeps expanding with the recent addition of fatty acids (e.g., de novo, mixed, preformed, polyunsaturated) to more standard components (e.g., fat, protein, MUN). Abnormal and unwanted trends can be detected in this data, which is generally done by producers and their advisors through visual inspection of components on a report.

Routine bulk tank analysis of milk fatty acids allows early detection of pattern changes which can provide early information on potential future problems. For example, a sudden and unplanned drop in de novo fatty acids concomitant with an increase in preformed fatty acids will often result in impaired rumen function, and eventually a drop in fat and protein yields, which will have an impact on milk revenue.

Trend anomaly detection on a report is tedious and requires skills, especially when simultaneously considering many complex variables such as specific fatty acids. Indeed, changes in trends often are not so obvious to detect by visual inspection, especially if they happen gradually over many days. In addition, there are many complex variables which are moving simultaneously in different directions. Also, experienced experts are not always available for consulting. One possible solution is to use robust analytics to provide insights to producers and their advisors, to help them reacting more rapidly and making more informed decisions.

This project aims to

1. assist in identification of anomalies in bulk tank milk components using basic statistical techniques;
2. inform/alert a producer and advisors that an abnormal trend is happening.

Complementarily, the use of a rule-based artificial intelligence (AI) approach is explored to help diagnosing and recognizing potential issues with respect to herd or rumen health management.

Our anomaly detection approach is based on the transformation of raw milk component values into features which we believe better represent an anomaly. These features, or indicators, are some basic statistical measures extracted from milk component values, and subsequently, a series of reports are produced for validation purposes on a daily basis. For running the aforementioned tasks, we developed a python package. However, the software should get enhanced to be capable of triggering messages and alerts based on extreme values in the new features. In parallel, we will continue to work with domain experts on the development of diagnostics rules.

Material and methods

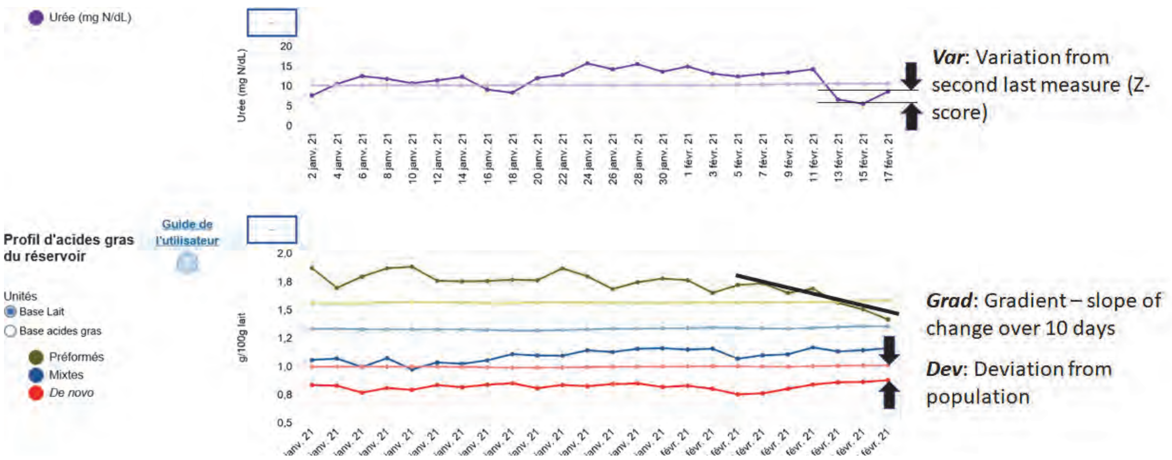
Data

The lab analysis is performed on a bulk tank milk sample of a typical herd every two days for extracting the profile of its components, including fat, protein, MUN, and fatty acids such as de novo, mixed, preformed, and polyunsaturated. In this project, there are seven time series that are considered for each herd. Having such data for about 1300 herds over 3 years, we calculate daily a moving average per breed for each component and use it as time series benchmark. All components are on a milk basis for this analysis (kg/hL for fat and protein, mg N/dL for MUN and g/100g milk for fatty acids).

Anomaly detection method

Our anomaly detection approach consists of two major steps: (1) Calculating statistical measures for all components of all herds, and (2) Ranking the herds, each component separately, within the same predominant breed. Specifically, in the first step, three measures are calculated for all components of each herd, as illustrated in Figure 1:

- *Deviation* is the difference between the component value and population benchmarks (see the chart lines of the de novo which show that herd values are lower than benchmarks). The advantage of using deviation measure is removing



Var: Variation from second last measure (Z-score)

Grad: Gradient – slope of change over 10 days

Dev: Deviation from population

Figure 1. Calculating deviation, variation, and gradient. Component profiles and their corresponding benchmarks are shown by dark and light colours, respectively.

trend or seasonality, and population reaction effects, such as thermal stress, in the original time series.

Variation is defined as the difference of the current component value from the second last measures (see the line chart of the MUN). We also calculate the z-score of the variation values for each herd to make results comparable with each other herd. Variations specify short-term changes.

- *Gradient* is defined as the rate of change in 10 days (see the line chart of the preformed which indicates a decreasing trend). Gradients represent trends or long-term changes.

It must be noted that the number of lags in variation and the number of days in gradient are considered as parameters which can be regulated based on expert opinion and on-going calibration.

After obtaining the measurements, we can rank the herds, each component separately, in the second phase. Rank, or more precisely, percentile rank, is a common metric for scoring that maps a quantity to a value between 0 and 100. For instance, if rank of the fat deviation for a herd is 20th, it means 20% of the herds are below that herd. Having three measures and seven milk components, we will end up to 42 indicators (21 raw values + 21 rank values) for each herd. It is worth mentioning that ranking is done within the same predominant breed. However, it also can be done without considering the breed, if the number of herds with specific predominant breed is scarce.

Transformation of the deviation, variation (z-score), and gradient raw values for each herd into percentile ranks has some advantages:

1. It is unitless.
2. We can easily select thresholds and extract data (e.g., top 10%).
3. It can be used as a proxy for probability, likelihood, or degree of confidence.

Diagnostic method

Our basic approach toward finding problematic herds is to assess component ranks in extreme cases, for each measure, separately, while user-defined thresholds define low and high extremes, where both are considered anomalies. Based on an inventive formula, we assign a confidence level to each measure/component of a herd, and consequently, a total confidence level to the herd, along with the number of confidence levels. For instance, consider gradient ranks for a herd are 4 in fat, 9 in protein, 1 in de novo, 0 in mixed, 89 in preformed, 98 in polyunsaturated, and 5 in MUN. If we set low and high thresholds to 10% and 90%, respectively, then the preformed value is filtered out and the confidence level for a component is calculated using the following relation:

$$conf = \begin{cases} 100 - rank & if\ rank \leq Low \\ rank & if\ rank \geq High \end{cases} \tag{1}$$

This leads to a confidence level greater than the high threshold for a component, i.e., 96 in fat, 91 in protein, 99 in de novo, 100 in mixed, 98 in polyunsaturated, and 95 MUN. Then, we report the average of the confidence levels over the six confidence levels as overall confidence level for the herd, which in this example equals to 97.

Although our basic approach recognizes problematic herds, it leaves the interpretation of anomalies and diagnosis to farm advisors or producers, which can be challenging. To assist producers and their advisors in diagnostics and selection of corrective actions, the use of a rule-based expert system approach is being explored. Currently, we are at the stage of eliciting, implementing, and validating diagnostic IF-THEN rules with the collaboration of domain experts and computer scientists. The rules are based on the anomaly indicators and ranks. In parallel, a prototype ontology-based diagnostic and action recommendation system was developed to explore the convenience of a symbolic AI approach in terms of experts' knowledge maintenance and automated reasoning capabilities,

Results and discussion

A python package was developed for trend anomalies and diagnostic. Using the package, daily anomaly reports for advisors can be produced. One report consists of the statistical measures by component and herd ranks. We also create symbolic rank reports to highlight the herds with the lowest or highest percentile scores, applying user-defined thresholds over the ranks. These results are intended to trigger alerts for producers and advisors in extreme cases.

Extreme value analysis

Currently, we have come up with two extreme cases which relate to the existence of severe problems in herds. However, these cases are not associated with a specific diagnostic. Interpretation of extreme cases remains with the observant:

- Extreme Case I: *De novo* is decreasing (i.e., the rank of the gradient is less than 10%) and preformed is increasing (i.e., the rank of the gradient is greater than 90%). See, for example, Figure 2. This case usually reflects a sudden body mobilization. Documented examples would include important stress affecting a great proportion of cows in the herd (e.g., diarrheal episode through the herd, feed delivery issue impacting ration composition for a few meals/hours), an expected but sudden change in ration composition (e.g., cows going out on pasture in early summer, addition of an oil-containing feed as home-roasted soybeans or potato chips) or an important proportion of cows freshly calved.
- Extreme Case II: *De novo* is decreasing, mixed decreasing, and preformed decreasing. See, for instance, Figure 3. Our experience shows that these cases are often related to an overall decrease in dry matter intake in the herd: Preformed fatty acids decrease because of the lower fatty acid intake directly; and de novo and mixed fatty acids decrease because of the decreased rumen precursor availability due to reduced rumen fermentation, therefore affecting mammary gland synthesis of these fatty acid groups. It is important to note that this change could also be due to an increase in milk yield (i.e., dilution effect on fatty acid concentrations on a milk basis), therefore a generally positive outcome for the producer.

It should be noted that we are developing a mechanism which not only recognizes severe problems, but also display attention messages on reports accordingly, and suitably sends alerts to the producers at a reasonable frequency (not too high to be unnecessarily bothering, and not too low to leave severe cases unrecognized). To find out if the above extreme cases can be considered as the base of the alert/attention system, we analysed the data over a 3-month period (February to April 2022), which contains 52,483 observations on 1271 herds. Table 1 shows a summary of the analysis: For instance, if only the trend of the de novo fatty acid is investigated, we see that about 90 percent of herds had at least one observation of negative trend in these

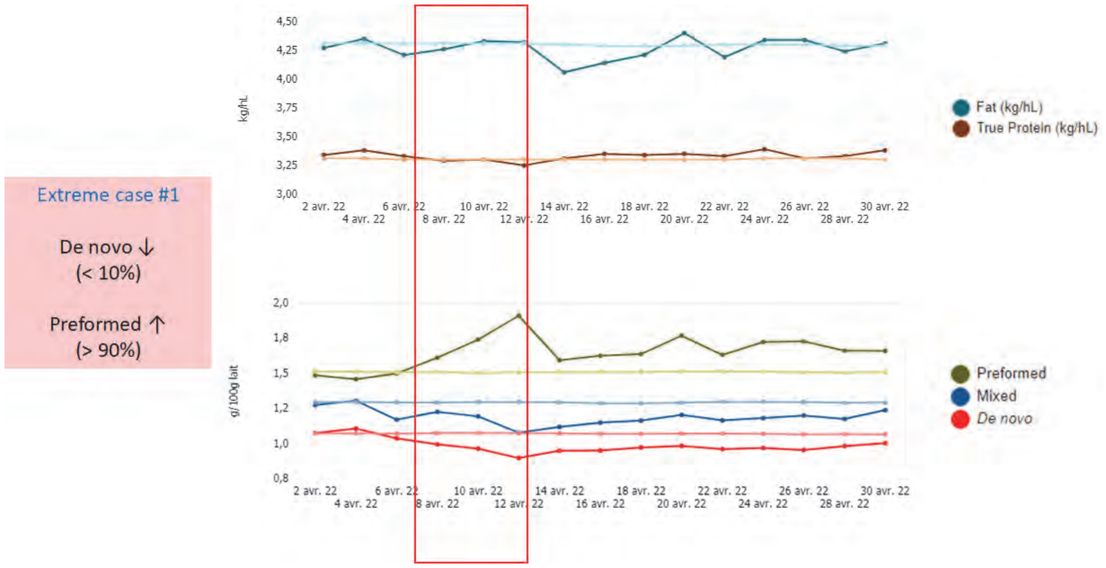


Figure 2. An example of the component profile of a herd in Extreme Case I.

Table 1. Analysis of extreme cases.

	% Observations	% Episodes	% Herds
All records	100%	100%	100%
De novo is decreasing	9.6%	5.5%	90.4%
Extreme Case I	1.7%	1.2%	36.%
Extreme Case II	0.7%	0.5%	18.6%

3 months. Or, equivalently, 9.6 percent of the records had an occurrence of decreased de novo (which corresponds to the fixed thresholds of 10%), among which 5.5 percent of records indicate episodes of decreased de novo, considering an episode is defined as a series of consecutive drops or increases. The average length of episodes in this case is equal to $\frac{9.6}{5.5} \approx 2$. Interestingly, the number of observations and episodes in both Extreme Case I and II is very low. Thus, both cases look qualified enough to trigger alerts along with proper messages on reports for producers at the beginning of an episode with an average length of 1.4 observations. Without triggering new alerts, attention messages can remain on reports during an episode.

With the help of domain experts in collaboration with computer scientists, we developed a rule-based system with seven diagnostic profiles and corresponding sets of corrective

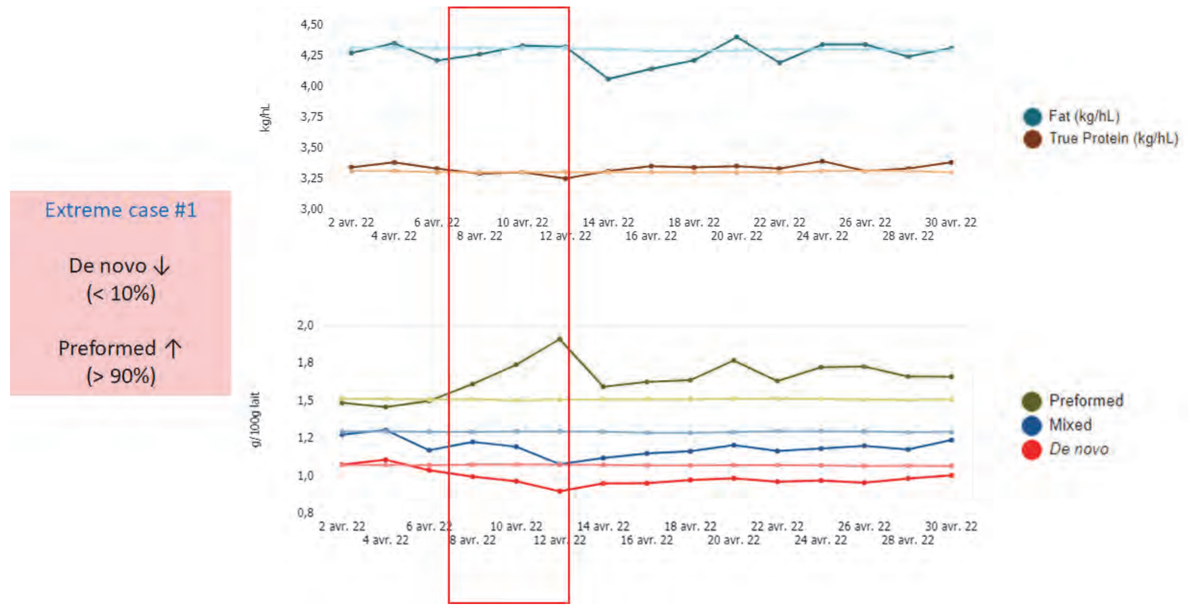


Figure 2. An example of the component profile of a herd in Extreme Case I.

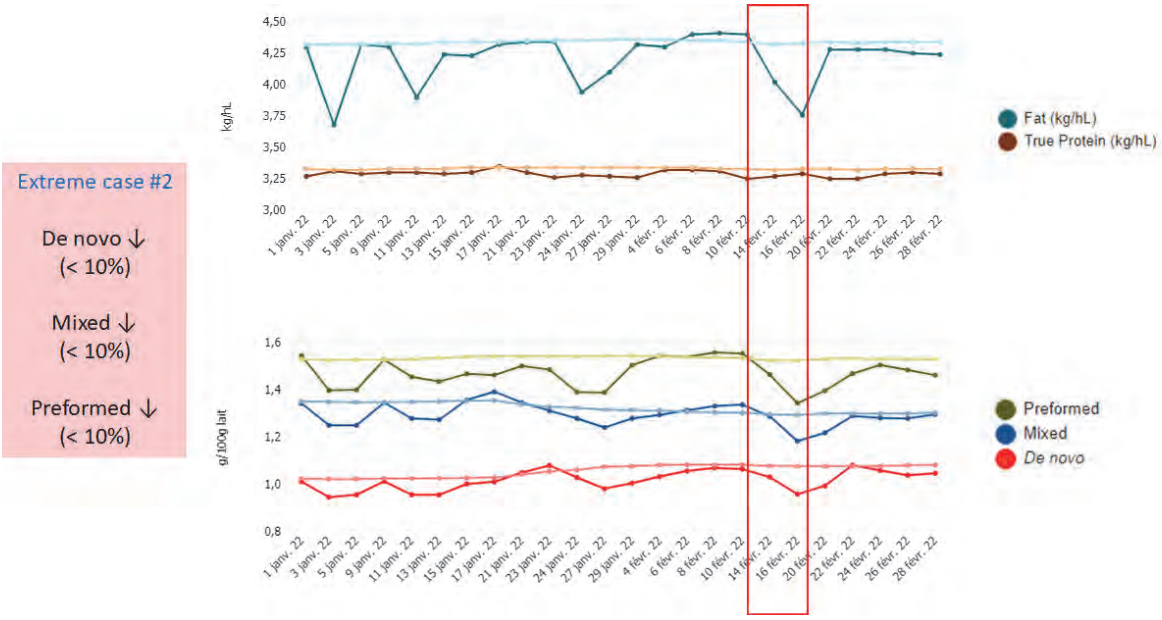


Figure 3. An example of the component profile of a herd in Extreme Case II.

Diagnostic rules

actions. Examples of profiles are risk of acidosis (Figure 4) and lack of consumption. One rule was developed for each profile, with each rule antecedent consisting of terms evaluating the value of selected deviations and gradients of milk components. A level of confidence in a diagnostic can be calculated through the ranks of the antecedents, similar to the previously presented extreme gradient rules. The rules were implemented in the anomaly detector program that runs every day. All rules are checked against all herds, and firing results are circulated for validation purposes.

Without surprises, this prototyping exercise has shown what has been known for years: there are challenges in the development, adjustment and maintenance of rule-based systems, and this approach leads to knowledge bases that tend to be static and limited in scope. For example, the possible corrective actions that were identified could individually be associated with different profiles, actions could be revised, and new actions could be frequently added, edited, or removed. Such knowledge could be embedded and handled more efficiently through an ontology, as we explored that avenue by implementing the seven rules and diagnostic profiles together with possible corrective action in a prototype. In this prototype, the ontology is stored in an OWL (Ontology-Web-Language) format and accessed by a python script. This script populates the ontology with the bulk tank data for each herd, which triggers an embedded reasoner and produces the corresponding diagnostics. Queries on the ontology can be done to retrieve possible corrective actions. Although, this prototype ontology was very limited, it could be expanded to make more complete recommendations, like suggestions about how to implement some corrective actions. In addition, such an ontology could be coupled to other existing ones, to expand the scope of recommendations of additional specific information.

In this project, we are developing a software application for recognizing anomalies in dairy herds, having bulk tank milk component profiles, such as fat, protein, MUN, and

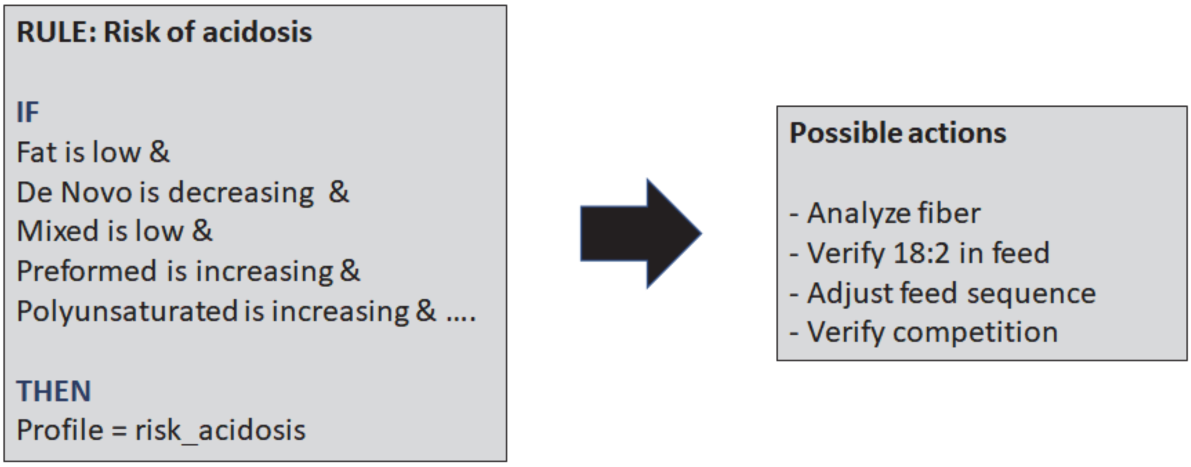


Figure 4. An example of a diagnostic rule and possible actions.

Conclusion

some specific fatty acids. This system will be equipped with an alert/attention system for producers and their advisors.

In the first phase of the project, we developed a python package to transform raw milk data to three statistical indicators (deviation, variation, gradient) that can represent anomalies, rank the herds based on the values of the indicators, and build daily reports which can be interpreted by experts to recognize anomalies. Based on such reports, we also investigated two extreme cases situations that can be related to severe problems (without implying a specific diagnosis) in herds, so that can be used for triggering alerts for producers and advisors. The extreme cases are identified from extreme fatty acids trends.

The second phase was focused on the development of a rule-based expert system. Currently, we are at the stage of building a knowledge base that consists of IF-THEN diagnostic rules for diagnosing issues such as risk of acidosis and lack of consumption. These rules are created based on indicator profiles obtained from the first phase, with the help of domain experts in collaboration with computer scientists. However, because of the challenges in updating and improving rule-based systems, we explored the use of ontology and symbolic artificial intelligence to develop a more complete diagnostic and recommendation system. Such an ontology can be expanded to cover additional dairy management areas in future.

Acknowledgement

The funding for this project was provided through the Government of Canada's Canadian Agricultural Strategic Priorities Program (CASPP).

PHENO3D: Phenotyping calves at weaning through automatic 3D image collection and instantaneous processing

M. Bruyas[1], C. Lecomte[1], C. Allain[2], A. Lebreton[2], L. Delattre[3], J.-M. Delouard[3], J.-M Gautier[2], O. Leudet[2], M. Brochard[2], P. Salvetti[1], and L. Griffon[4]

[1]ELIANCE, 149 rue de Bercy, 75012 Paris, France
[2]IDELE, 149 rue de Bercy, 75012 Paris, France
[3]3D Ouest, 5 Rue Louis de Broglie, 22300 Lannion, France
[4] Races de France, 149 rue de Bercy, 75012 Paris, France

Abstract

The Pheno3D project aims to bring the selection of beef cattle into high-throughput phenotyping by relying on the non-invasive technology of three-dimensional (3D) imaging. Thanks to artificial intelligence, the automation of 3D image analysis is now possible. Thus, from an existing 3D “scanner”, the PHENO3D project aims to develop an automated processing device (artificial intelligence) and in real time of 3D images to extract phenotypes used in the bovine meat sector (weight and scoring notes). To develop and validate the PHENO3D artificial intelligence, the project will first have to build a database comprising 3D images of the animals and their phenotypes (weight and scores) of approximately 500 images for each of the 10 involved breeds in beef recording. Artificial intelligence (Machine Learning, Deep Learning) will be optimized on its phenotype prediction performance but also on its ability to be embedded in a 3D scanner. The project will go beyond proof of concept by building a 3D imaging phenotyping service and its business models, and developing the skills needed for these use cases. To achieve its objectives, the PHENO3D project will rely on a network of actors covering the entire animal selection sector (RandD actors, performance control actors, breeding organisations). This use case in beef cattle will be a first step towards the deployment of high-throughput phenotyping by 3D imaging.

Keywords: 3D imaging, phenotype, performance, cattle, weighing, scoring, genetic.

Introduction

In France and in the beef cattle sector, genetic improvement is based on performance monitoring (also called phenotyping) of a large population of animals (Griffon *et al*., 2017). Phenotyping is based in part on a network of farmers affiliated to breeding organisations. For beef calves a first phenotyping is done around weaning. This collection of performances is carried out by advisers from Eliance network or from breeding organisations on farm. It includes animal weighing and scoring morphological traits. The morphological assessment comprises 19 scores on muscular and skeletal development or functional traits (see Figure 1). The compilation of performances allows, based on the phenotypic characterization of the breeders, the reasoned choice of mattings to meet the objectives of breed evolution. Even if the scoring by visual scores remains the most reliable, fast and less expensive technique compared to taking manual measurements on animals, the beef sector has expressed its need

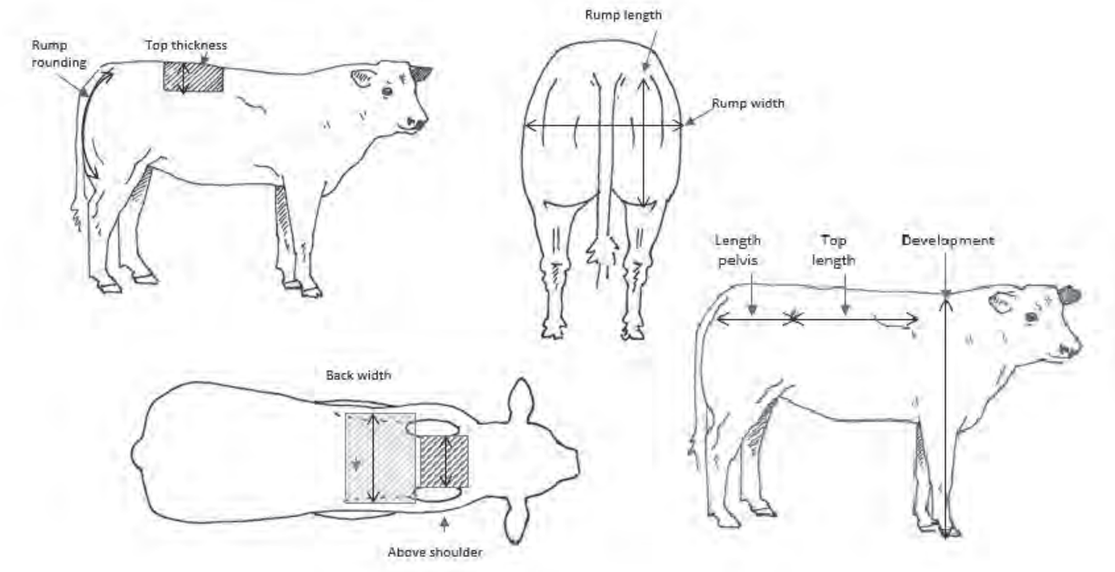


Figure 1. Syntheses of the positions of muscular development on the left and skeletal development on the right of beef cattle at weaning.

for an automatized digital solution to lower the very significant training costs and to reduce the human subjectivity effect on measurements.

The project partners have significant background and experience in the field of 3D imaging. At the scientific level, IDELE in partnership with INRAE and 3D Ouest led the morpho3D project. The objective was to estimate the weight and morphological measurements of dairy cows from 3D images of the entire animal (Le Cozler *et al*., 2019a; Le Cozler *et al*., 2019b). Another project involving the same partners aimed to obtain the body condition score of dairy cows from the cows' back 3D images (Fischer *et al*., 2015).

At the field level, FCEL, IDELE and Ingenera have developed a system (BodyMat M) to measure dairy cows BCS using a cane equipped with a 3D camera. Another project (BodyMat X) carried out with the same partners aimed to carry out morphological scoring of beef cattle using a scanner equipped with 3D cameras. Unfortunately, these initiatives failed to be industrialised with success because of difficulties with the industrial partner.

Finally, 3D imaging is a technology that has been proven to be the basis of prediction models for both simple metric measurements such as weight (Le Cozler *et al*., 2019a; Le Cozler *et al*., 2019b) or more qualitative notations. complex such as the body condition score (Fischer *et al*., 2015). If the collection of 3D images has been the subject of convincing proofs of concept, the processing of these images is not yet automatic and remains time-consuming. However, the application of finer artificial intelligence methods (such as simple or deep neural networks) on these images makes it possible to automate 3D image processing and analysis.

Considering the real needs and the past scientific, technical, and commercial experiences, we decided to run our own project from the development to the commercial phase.

The objectives of the Pheno3D project are to:

1. define the specification of the 3D scanner adapted to young beef cattle and for a mobile use;
2. define the best logistical arrangements for the installation and the passage of animals;
3. validate the ability of the system to gather good quality 3D images of entire beef cattle in movements:
4. develop and validate AI algorithms to estimate automatically morphological scores and weights of young beef cattle.

General project's road map

Main objectives

The tool must produce one or several 3D images to reconstitute an entire animal as a whole or partly. From this reconstruction, features will be extracted automatically. From the features or directly from the images, models and algorithm will estimate:

- The morphological scores for the traits described in figure 1.
- The weight with a prediction error or 3% max.

Device specifications

The device must be portable or transportable (and in this case be installed and ready to collect data in 30' and uninstalled in 20'). The dimensions of the device should not exceed 3m wide and 3.5m high which corresponds to the dimensions necessary to obtain qualitive 3D images (the smaller the distance between the animal and the sensor, the better the image quality) while avoiding scaring the animal to go inside. It must be transportable in a vehicle usually used by technicians and advisors. It will also have to be light (15kg of simultaneous load maximum) so that a single person can handle it. Indeed, it will be used by technicians who will move it from one farm to another.

The device must be able to produce a 3D image from a moving animal that will be guided within the device (using cables or fences). Indeed, the idea is to capture the data during the passage of animals (walking/running), one after the other, in a containment corridor.

The 3D cameras must be able to take a picture indoor or outdoor so they mustn't be too much sensitive to sunlight. The activation of the device can be done from human intervention. It will then have to be able to connect with an electronic animal identification device for farms where the animals are equipped with electronic loops.

The device will have to provide the expected score estimates in real time without internet access. The data must be exportable to another system.

The technical characteristics of the tool will be refined in the various specifications written throughout the project.

Targeted animals

The system will be used on beef calves at weaning (currently between 5 and 9 months). The algorithms will be developed for the French fallowing breeds: Charolaise and Limousine to start the project (being the 2 most developed breeds in France) then Aubrac, Bazadaise, Blanc Bleu, Blonde d'Aquitaine, Gasconne, Parthenaise, Rouge des Près and Salers will be integrated.

Conditions of use and measurement

The animals will have to pass through the device and the images will be taken on the moving cattle. The animals will therefore have to pass in motion, without running, without jumping, and perpendicular to the gantry to obtain clear images by the 5 pairs of sensors present on the device. To achieve this goal, field tests will be organized to optimize the speed and position of the animals. Several configurations will be tested to obtain the best results in order to validate the tool.

Device test and linear measures validation

The objectives will be to validate the ergonomics and the robustness of the device, the quality of the recorded images, the speediness of the images collection, in different conditions. This action must make it possible to adjust the prototype if necessary.

To validate the accuracy of the device, this action will aim to compare the measurements estimated by the prototype with reference measurements (like withers height, hips width, chest depth etc., see Figure 2) done manually on the animals and under different conditions.

This action will make it possible to estimate the prediction errors, repeatability and reproducibility of the 3D image collection and will validate the ability of the device to take a representative image of the real cattle. The exigence criteria for the validation are presented in table 1.

Development and validation of AI algorithms to estimate morphological scores and weight

The automatic processing of 3D images will be a revolution since in this project we want to achieve high-throughput phenotyping. Previously, image processing was largely manual, took much time and was not compatible with high-throughput phenotyping. In a performance monitoring use case, it is necessary to have a direct information in the 3D image quality to eventually make the animal pass a second time under the 3D scanner device. Having an automatic processing of the images, will allow advisers and

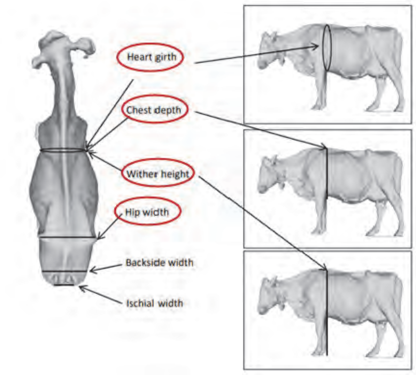


Figure 2. Morphological parameters tested to assess the quality of the image and measurements (in red).

Table 1. Exigency criteria for the device's validation.

Criteria	Values
Image quality	We will use the notion of completeness of the image. For this, we will look at the part of the animal that fits in a virtual box going from the hocks to the withers and from the tip of the buttocks to the end of the neck. This image should be complete. That is to say that if all the measurements defined above are feasible, the image will be considered complete and usable. 80% of the images must be usable on the first pass and 95% after 2 passes through the device.
Measurement precision	Correlation coefficient (Pearson) between manual measurements and measurements estimated on the 3D image > 0.7
Repeatability and reproducibility	Coefficients of variation of repeatability and reproducibility < 4%

farmers to have access to weight and scores quickly and discuss the results to adjust the monitoring of the herd in a unique session. Deep learning or machine learning models will be used to predict the weight and each morphological scores that are part of the scoring. The models will be built on a large dataset (more than 500 animals scanned, weighted and scored) built on farm to calibrate and then validate every models. The building of the dataset will be done on several farms during the current performance monitoring done by trained morphological scorers to ensure to capture the whole variability of the breed and thus having reproducible models.

The development of models by breed will be an important step since each breed has its morphological specificity and its own variability so the scoring has to be done based on breed standards. Because of this intra-breed specificities, models will have to take a breed effect or to have a model by traits to predict. The idea is to see if models can be efficiently adapted to one breed to another and if breed with a lower population (and probably a lower dataset) can benefit of the datasets from other breeds.

Prototype adjustment and development of the final version

Possible adjustments to be used routinely and safely (robustness, sensor protection, rugged boxes, data transmission and storage protocols, user manuals, transport cases, etc.) will be necessary as well as optimizations to limit construction costs. Development of the complete application for the end user will be done and will integrate their needs and requests. The final prototype integrating hardware and software adjustments will lead to the pre-industrial version of the device.

First results and achievements

Development of the first prototype

A first prototype was developed and tested on the field. The picture 3 below shows the main features and characteristics of the device.

Figure 3. On the right, the plan of the prototype; on the left, the prototype tested on the field; the prototype includes 5 pairs of cameras; the height is 2.6 m; width is 3 m

A few images were taken on the first trial. The first version tested only had two 3D sensors on the top of the scanner. A sample of the images taken are presented in figure 4 below. Images processing will allow a smoothing of the surface. Will all the sensors, the 3D images quality seems promising for extracting phenotypes.

Figure 3. On the right, the plan of the prototype; on the left, the prototype tested on the field; the prototype includes 5 pairs of cameras; the height is 2.6 m; width is 3 m

Figure 4. Depth and 3D images taken from the first prototype.

Optimization of the passage of animals in the scanner

The objective was to define scanning organizations to have animals that:

- Will pass under the device walking or trotting in the most standard position.
- Will not deteriorate the 3D sensors

On-farm tests allow us to define 2 organizations for scanning the animals. One, where the device is directly placed at the exit of a weighting scale without any additional fences or restraining elements (figure 5). The animals are managed in the way defined on the farm for weighting. The other requires two empty and adjacent pens of a barn or of a contention system. The device is placed in the first pen with additional fences that create a “funnel” from the width of the pen to a gate of 1.5 m (figure 6). Then the animal pass under the device and an operator guides it in the second pen where the animal will wait the end of the operation

Figure 5. Organisation of the scanning at the exit of the weighting scale.

Figure 6. Organisation of the scanning in a pen. White fences are additional fences added to form a funnel.

Summary of the road maps

Below the provisional schedule of the PHENO3D project is described:

- Spring to fall 2022: prototype testing phase in various situations to find the best way to pass animals in the device and obtain the best results
- Fall 2022: validation of the prototype: having good results passing animals
- Fall 2022 to spring 2023: data collection with several prototype to have lots of data by breed and to create algorithms.
- Spring 2023: validation of algorithms by breeding organisations. Each organisation will validate the algorithms concerning its own breed.
- During 2023: industrialization of the tool. 3D Ouest or another IOT manufacturer will industrialize the tool. We have estimated a need for more than 100 tools for France
- During 2024: routine use in the field

Conclusion

Automated morphological scoring based on 3D imaging will meet the sector's demand for an alternative to the current costly method: visual observation by technician scorer. The tool will be designed to obtain scoring and weighing data in a single scan of the animal. Today the weighing is carried out 2 to 3 times before weaning, the score only once. By adapting the system on the weighing sites, we will be able to collect more data on the morphology of the animal, at different ages, and new phenotypes like volume and body surface of the animal. This could improve selection schemes and advice services on sorting animals on farms.

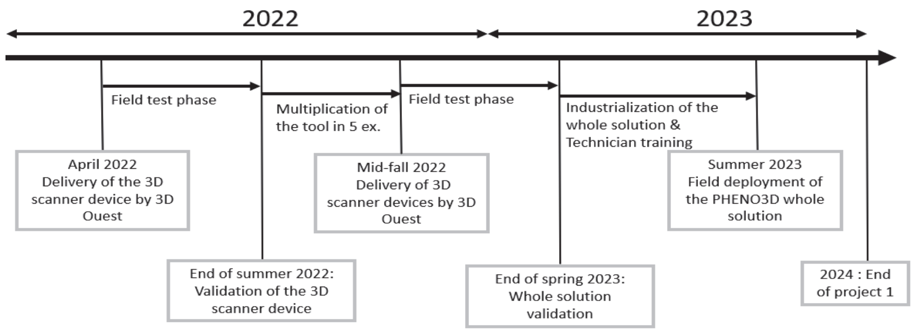


Figure 7. Project road map.

We also want to export the device internationally for weighing cattle without restraint, as the French scoring system is not used internationally.

In the medium term, artificial intelligence developed in the project will facilitate research on new phenotypes for beef cattle. Exporting the technology and the artificial intelligence developed in PHENO3D to animal other ages, other types of animals, or other sectors will make it possible to make significant advances in the collection and development of phenotypes for multi-performance breeding.

References

L. Griffon, P. Boulesteix, A. Delpeuch, Armelle Gion, J. Guerrier, *et al*. La sélection génétique des races bovines allaitantes en France: Un dispositif et des outils innovants au service des filières viande. INRA Productions Animales, Paris: INRA, 2017, 30 (2), pp.107-124. hal-01607593

Le Cozler, Y., Allain, C., Caillot, A., Delouard, J. M., Delattre, L., Luginbuhl, T., and Faverdin, P. (2019a). High-precision scanning system for complete 3D cow body shape imaging and analysis of morphological traits. Computers and electronics in Agriculture, 157, 447-453

Le Cozler, Y., Allain, C., Xavier, C., Depuille, L., Caillot, A., Delouard, J. M., and Faverdin, P. (2019b). Volume and surface area of Holstein dairy cows calculated from complete 3D shapes acquired using a high-precision scanning system: Interest for body weight estimation. Computers and Electronics in Agriculture, 165, 104977.

Fischer, A., Luginbühl, T., Delattre, L., Delouard, J. M., and Faverdin, P. (2015). Rear shape in 3 dimensions summarized by principal component analysis is a good predictor of body condition score in Holstein dairy cows. Journal of dairy science, 98(7), 4465-4476.

Challenges and opportunities in the use of milk MIR spectra: experiences learned from the OptiMIR and HappyMoo projects and beyond

N. Gengler[1], J. Leblois[2], OptiMIR Consortium[3] and HappyMoo Consortium[3]

[1]Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium
Corresponding Author: nicolas.gengler@uliege.be
[2]European Milk Recording, 5590 Ciney, Belgium
[3]Lead partner: Walloon Breeders Association, 5590 Ciney, Belgium

Abstract

Milk mid-infrared (MIR) spectral data is a rich source of information. Even if the technology has been known for decades, exploring its capacities beyond the estimation of major milk components is rather new. Experiences have shown the necessity to develop high levels of cooperation between industry and science but also across countries. The former North-West European Interreg project OptiMIR created an innovative European network of Milk Recording Organizations (MROs) that shared existing tools and approach of services to farmers, harmonized the way to record key data and allowed the access to the latest technology in getting information out of milk analysis. The tools developed by OptiMIR also allowed MROs easier development of implementations of MIR-based predictions and tools. One major achievement was to develop a method allowing the standardization of different MIR spectrometers across laboratories, and brands of apparatus. This standardization is now available to industry through an association of European MROs called European Milk Recording (EMR). Amongst the most important current issues in dairy herd improvement is the research of technologies for health and welfare monitoring. The current North-West European Interreg project HappyMoo develops novel strategies, methods, and tools to address this taking up the linked challenges but also feeding on opportunities in the use of MIR spectra. Amongst the five freedoms of animal welfare, we are addressing especially three, the absence of disease, hunger and stress. The ongoing research shows the importance of clear final objectives, precise trait definitions and collaborative work of industry and science. Among the topics that will be exemplified are the difficulties to combine reference data across systems (scales x implementation) for lameness, the problems to use sensor data provided in the field for BCS, the issue of synchronising reference and MIR data and the design of experiences to generate the needed reference data. The obtention of correct, variable, and validated reference data is one of the major challenges. Particularly considering more the biological background, but also concepts like deep phenotyping and molecular phenotypes will play certainly also a role in the future. Moreover, MIR-based phenotypes are always linked to the availability of the, at the most monthly, MIR data and innovative ideas are needed to go beyond this limitation. Advances in computational strategies will be needed as the use of pre-defined calibrations should be replaced by more advanced cloud-based learning and decision-making algorithms integrating sensors or other on-farm technologies.

Introduction

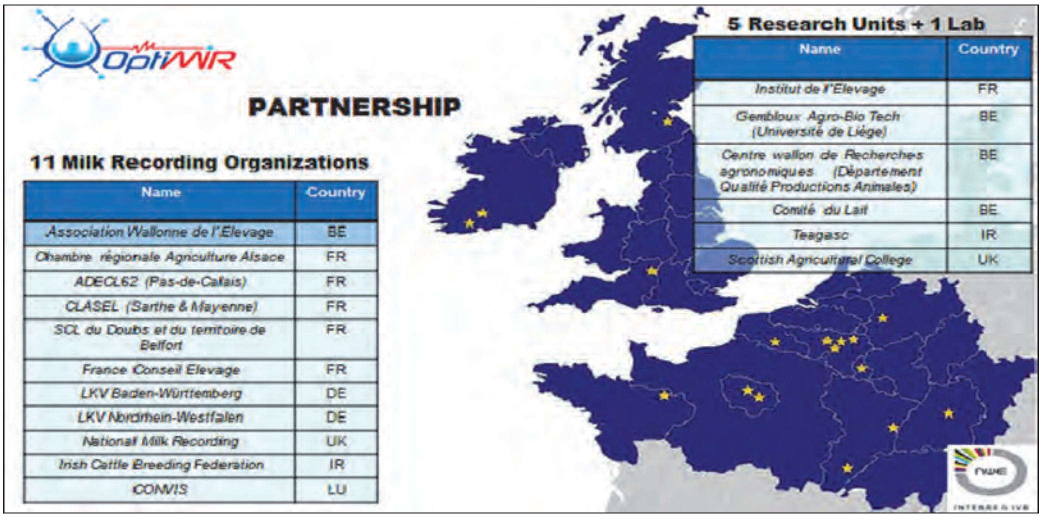
Importance of fine milk composition for dairy cow management and breeding was already highlighted by several previous reviews (e.g., Gengler *et al.,* 2016) and the use of milk mid-infrared (MIR), often also called FT-MIR, for Fourier-Transform MIR as currently predominant implementation, spectrometry (e.g., De Marchi *et al*., 2014) has been recognized in this context. The technology has evolved for decades but until rather recently in milk analysis its use was limited to major milk components. This document draws major conclusions on the challenges and opportunities in the use of milk MIR spectra in the context of management and breeding. Emphasis will be given to the experiences learned from the OptiMIR and HappyMoo projects and additional aspects beyond these will be highlighted.

Context

Large-scale phenotyping using milk MIR spectra was extensively developed during the last years. This phenotyping has become an important source of improvement in herd management, animal genetic evaluation, and control of quality of milk and subsequently milk products. But as highlighted by Grelet *et al.* (2021) key factors affect the quality of prediction. As a matter of fact, a priory the milk MIR spectra are very similar to marker genomic data and needs to be “calibrated” to make it a posteriori useful. Currently extensive research is ongoing to advance in the understanding of the requirements to achieve high quality prediction equation. Research illustrated the need to address several issues as the precise definition of the target trait (reference phenotype) and its nature, the excellent control of the variability in reference and MIR data, both in the calibration datasets but also the application datasets. Also, some merely technical issues were highlight as the potential difficulty to get access to raw spectra and the need to harmonize spectra coming from different brands models and machines, but also improve stability over time. All these experiences have shown the necessity to develop high levels of cooperation between industry and science but also across countries.

OptiMIR and HappyMoo

From 2009 to 2015 supported by 3.7 M€ European Union funding through the North-West European Interreg scheme the OptiMIR project (https://keep.eu/projects/6989/OptiMIR-new-tools-for-a-more-EN/) joint 11 milk recording organizations, 1 lab and 5 research units creating an innovative European network of Milk Recording Organizations (MROs). Figure 1 gives the partnership but also the geographical distribution across 6 countries. These MROs shared existing tools and approach of services to farmers, harmonized the way to record key data and allowed the access to the latest technology in getting information out of milk analysis. Very relevant was the joint acquisition of reference data covering the variability across the different production systems but also the joint developments of functional specifications. This allowed MROs much easier development of implementations of MIR-based predictions and tools. From a research point of view several equations were improved, new equations initiated. One major achievement was the development of a method allowing the standardization of different MIR spectrometers across laboratories, and brands of apparatus (Grelet *et al.,* 2015). This standardization is now available to industry through an association of European MROs called European Milk Recording (EMR) (https://www.milkrecording.eu/).



PARTNERSHIP

11 Milk Recording Organizations

Name	Country
Association Wallonne de l'Elevage	BE
Chambre régionale Agriculture Alsace	FR
ADECL62 (Pas-de-Calais)	FR
CLASEL (Sarthe & Mayenne)	FR
SCL du Doubs et du territoire de Belfort	FR
France Conseil Elevage	FR
LKV Baden-Württemberg	DE
LKV Nordrhein-Westfalen	DE
National Milk Recording	UK
Irish Cattle Breeding Federation	IR
CONVIS	LU

5 Research Units + 1 Lab

Name	Country
Institut de l'Elevage	FR
Gembloux Agro-Bio Tech (Université de Liége)	BE
Centre wallon de Recherches agronomiques (Département Qualité Productions Animales)	BE
Comité du Lait	BE
Teagasc	IR
Scottish Agricultural College	UK

Figure 1. OptiMIR partnership and geographical distribution.

Based on this success and in order to address one of the most important current issues in dairy herd improvement, the research of technologies for health and welfare monitoring, in 2018 several of the groups involved in OptiMIR (https://keep.eu/projects/21152/Delivering-NWE-dairy-farmer-EN/) but also new partners, joint forces in 2018 for a four year project supported by 2.3 M€ European Union funding through the North-West European Interreg scheme (Figure 2). OptiMIR develops novel strategies, methods, and tools to address this taking up the linked challenges but also feeding on opportunities generated by OptiMIR.

Amongst the five freedoms of animal welfare, especially three, the absence of disease, hunger and stress are addressed by HappyMoo.

Partnership

- 8 milk recording organisations
- 3 research centres
- 1 IT centre
- 1 milk pilot lab

→ **12 partners in total**
→ **In 7 countries:** Belgium, France, Germany, Ireland, Luxemburg, Switzerland, United Kingdom

Figure 2. HappyMoo partnership and geographical distribution.

Lessons learned

The ongoing research shows the importance of clear final objectives, precise trait definitions and collaborative work of industry and science. Among the precise issues that were found were the difficulties to combine reference data across systems. This was recently exemplified for lameness illustrating the fact that harmonization is not only an issue of scales but also their implementation and use are critical. Similarly, in the context of studying BCS provided by sensors, problems to use this data, the issue of synchronising reference and MIR data and the design of experiences to generate the needed reference data is an important issue. The obtention of correct, variable, harmonized and validated reference data is one of the major challenges. Here the role of ICAR as harmonizing body comes into focus. However, if for good reasons, reference data reflects differences in traits, breeds, circumstances, the question that needs to be answered if a general prediction equation can be achieved, or even ii the preferable option. Equations adapted to specific situations could be an alternative.

Potential future challenges and opportunities

A bottleneck that MIR-based phenotypes cannot avoid is the precise definition and recording of needed high quality reference phenotypes. Novel concepts like deep phenotyping and molecular phenotypes adding more biological background will play certainly a role in the future. Moreover, MIR-based phenotypes are always linked to the availability of the, at the most monthly, MIR data and innovative ideas are needed to go beyond this limitation. Advances in computational strategies will be needed as the use of pre-defined calibrations should be replaced by more advanced cloud-based learning and decision-making algorithms integrating sensors or other on-farm technologies. Finally, all developments will have to be integrated in the development of technical, scientific, and regulatory frame works, obviously a field were ICAR but also IDF potentially through ExtraMIR (Extra value from -smart use of- MIR spectra), a new joint action team of both organisations, will be important players.

Conclusion

Uses of MIR spectra in the dairy sector are increasing. International collaboration as exemplified by OptiMIR and HappyMoo with the objectives to offer concrete solutions and tools, both for MROs but also at the end to the farmer, was recognized as a promising strategy. Despite these experiences, current challenges, and opportunities in the use of milk MIR spectra subsist and new ones appear.

Acknowledgement

The authors also acknowledge the important support they received in the field of MIR research throughout the Futurospectre partnership, the “Service Public de Wallonie” (SPW – ARNE, Belgium), the National Fund for Scientific Research (Brussels, Belgium), the EMR EEIG and different European Projects (FP7: Robustmilk - Grant agreement n° 211708, GreenhouseMilk – Grant agreement n° 238562, GplusE - Grant agreement n° 613689, and H2020: The SmartCow - grant agreement n° 730924, INTERREG NWE: OptiMIR NWE190G and HappyMoo NWE730) supported this research. The content of the publication reflects only the view of the authors; the European Community is not liable for any use that may be made of the information contained in this publication.

References

De Marchi, M., V. Toffanin, M. Cassandro and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits1. Journal of Dairy Science 97:1171–1186. doi:10.3168/jds.2013-6799.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M.-L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows1. Journal of Dairy Science 99:4071–4079. doi:10.3168/jds.2015-10140.

Grelet, C., P. Dardenne, H. Soyeurt, J.A. Fernandez, A. Vanlierde, F. Stevens, N. Gengler, and F. Dehareng. 2021. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. Methods 186:97–111. doi:10.1016/j.ymeth.2020.07.012.

Grelet, C., J.A. Fernández Pierna, P. Dardenne, V. Baeten and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. Journal of Dairy Science 98:2150–2160. doi:10.3168/jds.2014-8764.

An equation based on MIR spectra to explore the genetic determinism of spontaneous lipolysis in dairy cows

M. Gelé[1], S. Meurisse[1], P. Trossat[2], A. Oudotte[2], D. Boichard[3], L. Bernard[4], C. Hurtaud[5], A. Barbat[3] and C. Cebo[3]

[1]Institut de l'Elevage, 149 rue de Bercy, 75595 Paris cedex 12, France
Corresponding Author: marine.gele@idele.fr
[2]Actalia-Cécalait, Rue de Versailles, 39800 Poligny, France
[3]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France
[4]INRAE, Université Clermont Auvergne, VetAgroSup, UMR Herbivores, 63122 Saint-Genès-Champanelle, France
[5]PEGASE, INRAE, Institut Agro Rennes-Angers, 35590, Saint-Gilles, France

Abstract

Mid infrared (MIR) spectra have been used since the late 2000s to phenotype new traits at large scale through prediction equations. It has been used in France since 2012 to quantify free fatty acids on herd milks in the framework of quality-based milk payment system. In addition to the loss of income, a high lipolysis level leads to a degradation of organoleptic (rancid taste) and technological (processing inability) properties of milk. Monitoring lipolysis is required to maintain the quality of milk and dairy products. Lipolysis depends on many factors: genetics, farming practices, milking equipment, transformation process. Regarding genetics, Vanbergue suggested in 2017 the existence of a genetic susceptibility to spontaneous lipolysis.

The present study aims to investigate this issue further, by studying the genetic determinism of spontaneous lipolysis in a larger number of cows. This work was carried out within the framework of the LIPOMEC project which aims to better understand the molecular mechanisms controlling the degradation of milk fat in dairy species. Studying the genetic determinism of lipolysis requires the phenotyping of many cows. As the equation initially developed on herd milk was not fully appropriate for individual milks (range, precision), a new prediction equation was estimated from the MIR spectra of individual milks. For this purpose, 432 milk samples were collected in 4 experimental farms in 2018 (approximately 40 cows per farm sampled 2 or 3 times a year) to maximize the variability of breeds (Holstein, Normande, Montbéliarde, Jersey) and diets.

A joint analysis of lipolysis according to ISO/TS 22113 standard (BDI method) and by MIR spectrometry was carried out on each sample. Lipolysis measured by BDI method averaged 0.53 mmol/100 g fat (sd=0.41 mmol/100 g fat). The equation was developed by Partial Least Square regression after LOG transformation. Its coefficient of determination R² reached 0.72, with a residual standard deviation $S_{y,x}$ of 0.19 mmol/100 g of fat. The equation was then applied to obtain phenotypes on more than 300,000 MIR milk spectra from Holstein, Normande and Montbéliarde breeds. Genetic parameters were estimated using a repeatability animal model. Heritability and repeatability estimates were moderate in both Normande and Holstein breeds but higher in Montbéliarde breed. This work opens the opportunity to new uses of MIR spectra to improve the control of lipolysis in farm, by a closer management of the herd, or even by a selection plan. The LIPOMEC project was funded both by APIS-GENE and the French National Agency.

Keywords: Lipolysis, dairy cattle, mid infrared spectroscopy, cow genetics.

Milk lipolysis is the breakdown of milk fat by an enzyme, the lipoprotein lipase, resulting in the release of free fatty acids in milk. This may lead to the development of a rancid flavor in milk considered unacceptable by the consumer beyond a certain threshold. Moreover, the presence of partial glycerides in milk interferes with technological processes in the industrial processing of dairy foods. In some French regions, lipolysis is therefore part of the criteria of the quality-based milk payment system, which leads to penalties when the level of lipolysis is above national standards (0.89 mEq/100g of fat).

Since 2012, lipolysis levels in milk are routinely measured in France by inter-professional milk analysis laboratories accredited under the milk quality payment scheme according to an instrumental method based on mid-infrared spectroscopy (MIR). However, this method has two drawbacks. First, the equation used was constructed from herd milks, and is not fully adapted to individual milk samples. Second, the anchor method used to check and adjust the calibration equation is the copper soap method, which is calibrated in the range from 0.4 to 1.2 meq/100 g fat, does not thus allow the identification of extreme values of lipolysis in milk.

Lipolysis results from a complex interplay between animal physiology, farming practices, milking equipment and technological process (Vanbergue, 2017; De Marchi et al. 2017). Recently, Vanbergue (2017) suggested the existence of a genetic susceptibility of dairy cows to lipolysis that needs to be confirmed.

The objective of this study was therefore to develop a new equation based on MIR spectra to predict the level of lipolysis in individual milk samples, with the goal of using this tool to phenotype the high number of milk samples required to further explore the genetic determinism of lipolysis. This work was carried out within the framework of the LIPOMEC project, the first integrative project aimed to better understand lipolysis mechanisms in dairy species and granted both by APIS-GENE and the French National Agency.

Material and methods

Four hundred and thirty-two milk samples have been collected to meet the above objectives. A joint analysis of lipolysis according to ISO/TS 22113 standard (BDI method) and by MIR spectrometry has been carried out on each sample.

Data collection

Collection of milk samples

Four hundred and thirty-two milk samples were collected from four experimental farms, located in several regions in France, between March and October 2018 (approximately 40 cows per farm sampled two or three times per year) to maximise variability in dairy breeds (Holstein, Normande, Montbéliarde, Jersey) and diets.

During sampling, vials containing 0.02% bronopol preservative (wt/vol) were filled to capacity (100 mL) to avoid “churning” of the milk that could damage fat globules and activate lipolysis during transport. After collection, milk samples were stored at 4°C to limit bacterial proliferation and lipase-associated activities. Milk samples were sent at 4°C to ACTALIA CECALAIT (39800 Poligny, France) for subsequent analyses.

Table 1 shows the distribution of these samples across farms and time periods.

Table 1. Distribution of samples collected between farms and sampling periods.

	Number of milk samples			
Experimental farm	**March/April 2018**	**June 2018**	**October 2018**	**Total**
Grignon	41	40	36	117
UE du Pin	40	36	40	116
UE Herbipôle	40	38	40	118
IE PL Le Rheu	41	0	40	81
Total	162	114	156	432

Recording of MIR spectra.

MIR spectra were recorded at ACTALIA CECALAIT using MilkoScan™ FT+ spectrometer (Foss, Hillerød, Denmark).

Measurement of lipolysis in milk.

Reference values for lipolysis in milk were determined using the ISO/TS 22113|IDF/RM 204 BDI (Bureau of Dairy Industry) method which specifies a method for determining the titratable acidity of milk fat. This analysis was carried out by ACTALIA CECALAIT within 36 hours of sampling. Lipolysis measured by BDI method averaged 0.53 mmol/100 g fat (sd=0.41 mmol/100 g fat).

Data processing

Development of the lipolysis prediction equation.

Four hundred and thirty individuals were retained after removal of two outliers for the development of the milk lipolysis prediction equation. The reference lipolysis values obtained by BDI method were log transformed. The equation was developed by partial least squares (PLS) regression using R software.

Calculation of genetic parameters of lipolysis

The lipolysis prediction equation was applied to 348,000 spectra from a database created in the PhenofinLait programme (2008-2013). After exclusion of outliers (very high values), 333,862 individuals were used for phenotypic description.

The estimation of genetic parameters was performed within breed using an animal model with repeated data over the first 3 lactations, i.e., 73,000 data for the Holstein breed, 71,000 for the Normande breed and 55,000 for the Montbéliarde breed. The following fixed effects were considered: herd x test day, analysis laboratory and time between sampling and analysis, month of calving within the season, age at calving in first lactation, stage of lactation (intra parity) and stage of pregnancy.

Results and discussion

Accuracy of the prediction equation of milk lipolysis

The developed milk lipolysis prediction equation has a coefficient of determination (R^2) of 0.72 and an error ($S_{y,x}$) of 0.19 meq/100 g fat. Figure 1 shows the prediction results obtained using this equation, compared with the reference values obtained using the BDI method.

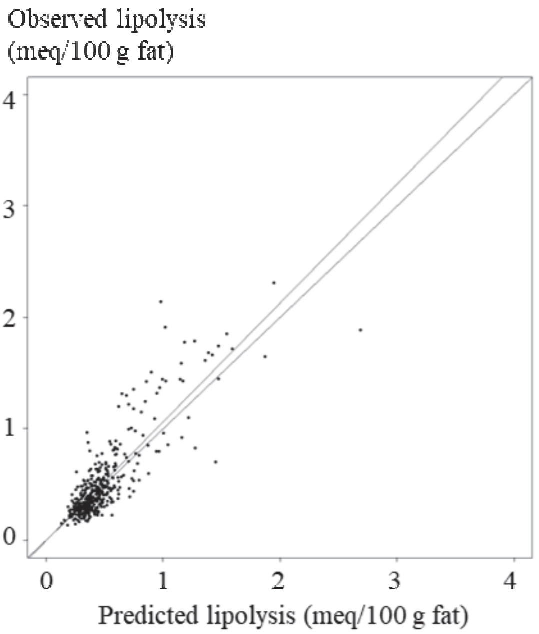


Figure 1. Lipolysis predicted by the developed equation compared with the values observed by BDI method.

The developed milk lipolysis prediction equation is sufficiently accurate to be immediately used for breeding advice or genetic research but is intended to be improved with the inclusion of new data and new samples for others purposes linked to the milk quality payment scheme.

Phenotypic variability of milk lipolysis

Data obtained from the PhenoFinlait database showed that milk lipolysis fluctuates during lactation with a more significant amplitude in first lactation.

The effect of breed was highlighted, with Normande displaying a lower lipolysis rate in milk than Montbéliarde, while Prim'Holstein being intermediate.

Genetic parameters of milk lipolysis

The heritability of the milk lipolysis trait is moderate to high, which makes it possible to consider genomic selection in the future. In addition, for all three breeds, several regions of the genome related to the trait 'milk lipolysis' have been identified, some of which are genes or gene regulatory regions whose involvement remains to be confirmed.

Conclusion

The newly developed equation can be used as a new precision breeding tool to monitor lipolysis in farms for a closer management of the herd. It opens the road to genetic selection to monitor lipolysis at the animal level.

Acknowledgement

This work received a financial support from APIS-GENE (https://www.apis-gene.com/). Authors are grateful to Philippe Lamberton (INRAE, IE PL, 35650 Le Rheu, France), Frédéric Launay (INRAE, Unité Expérimentale du Pin, 61310 Le Pin au Haras, France), Matthieu Bouchon (INRAE UE 1414 HERBIPOLE 15190 Marcenat, France) and Pierre-Henri Pomport (AgroParisTech-Ferme experimentale de Grignon, 78850 Thiverval Grignon, France) for their contributions to animal care and milk sampling.

References

De Marchi, M., Penasa, M., Cassandro, M., 2017. Comparison between automatic and conventional milking systems for milk coagulation properties and fatty acid composition in commercial dairy herds. Italian Journal of Animal Science 16(3): 363-370.

Vanbergue E., 2017. Les facteurs de variations de la lipolyse spontanée du lait de vache et mécanismes biochimiques associés. Thèse de doctorat, Renne, Agrocampus Ouest.

Predicting body condition score change in early lactating Irish dairy cows using milk mid infrared spectra

M. Frizzarin[1,2], I.C. Gormley,[2] D.P. Berry,[1] and S. McParland[1]

[1]Teagasc, AGRIC, Moorepark, Fermoy P61 P302, Co. Cork, Ireland
Corresponding Author: donagh.berry@teagasc.ie
[2]School of Mathematics and Statistics, UCD, Ireland

Abstract

Body condition score (BCS) is a subjective metric used worldwide to reflect the fat stored in mammals. BCS, as well as its change in early lactation, has been associated with productive, reproductive, and health traits. The objective of the present study was to predict, using different statistical techniques, BCS change (ΔBCS) in early lactation dairy cows using milk mid infrared (MIR) spectra. A total of 73,193 BCS records from 6,572 cows were collected from five research farms. To generate daily ΔBCS, splines with 6 knot points across days in milk (DIM) were fitted through individual test-day records of BCS. Daily BCS was interpolated from the splines and used to calculate daily ΔBCS. ΔBCSobservations were merged with MIR spectra recorded on the same week. Data in the first 120 DIM were retained. Three statistical methods were used to predict ΔBCS from the spectra; partial least squares regression (PLSR), generalized additive mixed model (GAMM), and neural networks (NN). Spectra and DIM were used as predictors in NN and PLSR, while the first 20 principal components of the spectra and a spline fitted through DIM were used as predictors in GAMM. Tuning parameters of PLSR were determined using 10 fold cross-validation. The NN model had two hidden layers and a Bayesian regularization applied to the input layer. To compare predictive ability across the approaches, the dataset was divided in 4 sub-datasets, and iteratively 3 sub-datasets were used to train the methods, while the remaining sub-dataset (the test dataset) was used to test the methods. Prediction accuracy was evaluated according to the root mean square error of the test dataset (RMSEV; here multiplied by 1000) and the correlation (r) between the actual and the predicted ΔBCS. The RMSEV and r obtained from the four test datasets were averaged. Body condition score change was predicted with an average RMSEV of 1.02 BCS units (SD=0.010) and r of 0.87 (SD=0.004) from NN; the SD of actual ΔBCS was $2.05*10^{-3}$ BCS units. Partial least squares regression performed better than GAMM with an average RMSEV of 1.06 (SD=0.010) and 1.10 (SD=0.010) BCS units, respectively, and an r of 0.86 (SD=0.004) and 0.84 (SD=0.004) for PLSR and GAMM, respectively. Results from the present study demonstrate the potential to use milk MIR spectra to predict ΔBCS, which can be used to support farm decisions and can be incorporated in dairy cow breeding programs.

Keywords: Mid-infrared spectroscopy, body condition score change, early lactation, machine learning.

Introduction

Body condition score (BCS) is a metric used worldwide to reflect fat stored in mammals (Roche *et al*., 2009). In dairy cows, BCS tends to be lost after parturition and it is generally restored before a new calving (Bauman and Carrie, 1980). Body condition score change is used on farms as an indicator of cow energy balance, and its trajectory in early lactation is generally used to support fertility decisions. In fact, BCS, as well as its change in early lactation, have been associated with fertility and health traits (Berry *et al*., 2007; Buckley *et al*., 2003; Roche *et al*., 2007). Buckley *et al*. (2003), for example, reported that the mean BCS at 60 and 100 days of lactation was positively associated with both submission for breeding in the first 21 days of the breeding season and the likelihood of pregnancy after 42 days of breeding. Butler and Smith (1989) documented the effect of the change of body condition score in early lactation on the rate of conception at first serving, reporting a conception rate of just 17% to first service in cows that lost more than 1 BCS unit (5-point scale) after calving, while cows that lost <1 BCS unit had a conception rate of 53% to first service. Body condition score, as well as its change (∆BCS), are therefore useful metrics for producers. Body condition score change can be calculated from multiple BCS records albeit these may not always be available.

Mid infrared spectroscopy of milk is a fast, cheap, and non-disruptive technique to generate spectra from milk samples (De March *et al*., 2013). The spectra are produced from the interaction between the light emitted from the spectrometer and the molecules within the milk and, utilizing suitable statistical techniques, it is possible to predict milk and animal traits from the spectra (Soyeurt *et al*., 2006; McParland *et al*., 2014). Milk MIR spectra are used to predict milk fatty acids with good accuracy (Soyeurt *et al*., 2006) and, as reported by Nogalski *et al*. (2012), cows that lost >1 BCS unit in early lactation had a different milk fatty acid profile compared to cows that lost < 1 BCS units. Thus there is a strong biological hypothesis as to why ∆BCS could potentially be predicted from milk MIR spectra. McParland *et al*. (2014) pioneered the research on predicting DBCS from milk spectra in dairy cows, reporting a correlation between the actual and the predicted ∆BCS ranging from 0.57 to 0.75 using different spectra combination and partial least squares regression (PLSR) as the prediction method.

The objective of the present study was to predict DBCS using morning milk spectra in early lactating dairy cows and to compare predictions from PLSR with those from generalized additive mixed model (GAMM) and neural networks (NN). Predicting ∆BCS from milk MIR spectra can routinely provide producers with DBCS estimates, which can be used in making reproductive decisions, as well as providing phenotypes for use in dairy cow breeding programs.

Materials and methods

Data and data editing

A total of 73,193 BCS records from 6,572 cows were recorded in 5 Teagasc research farms between 2015 and 2019. Body condition score was assessed by trained scorers using a 5-point scale with increment of 0.25 (Edmonson *et al*., 1989). BCS was recorded every 18 days, with a repeatability of BCS within a 7-day period being 0.69. Cows were from different parity orders (from parity 1 to 12) and different breeds (Holstein-Friesian, Jersey, Norwegian Red, as well as their crosses). Daily BCS for each lactation was calculated after fitting a cubic spline with 6 knot points at 20, 70, 120, 170, 220, and 270 days in milk (DIM) through individual test-day records of BCS with a covariance structure fitted among knot points (McParland *et al*., 2014). Daily ∆BCS was calculated from the fitted splines as the BCS on a given day minus the BCS of the previous day. To avoid potential problems with interpolation, only DBCS calculated on the day of an actual BCS records were retained. Lactations greater than 10 were deleted, and only DBCS records between DIM 5 and 120 were retained. Parities were grouped as first, second, third, fourth, and fifth or greater, and DBCS outliers were deleted for each parity group as DBCS records greater than 3 standard deviations (SD) from the mean ∆BCS of the respective parity group.

From the same cows and across the same time period, 423,532 individual milk spectra were collected from morning milking. Spectra outliers were deleted following the same procedure as described by Frizzarin *et al*. (2021a), and were standardized across time using the standardization coefficients provided as part of the European Milk Recording project ring testing program (Grelet *et al*., 2015; Grelet *et al*., 2017). Body condition score change records were merged with the closest spectra recorded within one week. The final dataset comprised 13,492 ∆BCS records across 2,489 lactations from 1,250 cows.

Prediction methods

All the analyses were conducted using the statistical software R (R core team). Three different prediction methods were used to predict ∆BCS: PLSR, GAMM, and NN. For the PLSR analyses, the a.m. spectra as well as a fourth order polynomial of DIM were used as predictor variables. The R package pls (Mevik *et al*., 2019) was used to develop the prediction equation. The number of PLSR factors were defined using 10 folds cross-validation (CV). For the GAMM analyses, the first 20 principal components of the a.m. spectra were used as linear predictor variables with DIM fitted as a spline; cow-lactation was accounted as repeated measures in the model. The R package gamm4 (Wood and Scheipl, 2020) was used to develop the prediction equation. Lastly, a Bayesian regularized NN was developed using the a.m. spectra and the DIM as predictor variables. The R package brnn (Perez Rodrigez and Gianola, 2020) was used to develop the prediction equation, and the default tuning parameters were chosen, which included two hidden layers and a Bayesian regularization to the input layer to improve generalizability.

Measures of prediction performance

To test the predictive performance of the methods on unseen data, the original dataset was divided into four sub-datasets, three of these were used to train the prediction equation (i.e., training dataset), and the fourth sub-dataset was used to test the equation (i.e., test dataset). This process was repeated until all the four sub-datasets were considered as test dataset once.

Different metrics were used to evaluate prediction performance, such as the root mean square error on the test dataset (RMSEV), the correlation between the predicted ∆BCS values in the test dataset and the actual ∆BCS values (r), the bias of the prediction, the slope, and the ratio of performance to interquartile distance (RPIQ). The bias corresponds to the average of the residuals, the slope corresponds to the slope of the line where the real ∆BCS values are plotted against the predicted ∆BCS values, and the RPIQ is the ratio between the interquartile range of the observed trait values and the RMSEV. All these metrics were calculated for each of the test datasets, and were subsequently averaged. The SD of the performance metrics across the four test datasets was also calculated and was considered as a reflection of robustness of the prediction method. The F-test was used to compare the RMSEV across prediction methods.

Results

The prediction performance of the different prediction methods on the test datasets used to predict ∆BCS are summarized in Table 1. Neural networks had the lowest RMSEV ($P < 0.05$) as well as the greatest r and RPIQ; thus, for this dataset, NN was the prediction method which predicted ∆BCS with greatest accuracy. Partial least squares regression had a slightly greater ($P < 0.05$) RMSEV compared to NN. Generalized additive mixed model was the prediction method with the poorest prediction performance ($P < 0.05$). For all prediction methods, the SD of the RMSEV and r was low, indicating that the prediction performance were robust across all the test datasets.

In Figure 1 is presented the average actual ∆BCS across DIM, as well as the average predicted values from PLSR, GAMM, and NN. Both GAMM and NN predictions perfectly followed the actual ∆BCS trend, while PLSR slightly over predicted the actual ∆BCS between week 3 and week 8 of lactation, and slightly under predicted actual ∆BCS between week 8 and week 12 of lactation.

Table 1. Prediction performance[1] on the test dataset of partial least squares regression (PLSR), generalized additive mixed model (GAMM), and neural network (NN) to predict body condition score change across 120 days in milk (DIM).

Method	RMSEV^{2,3} (SD^{2,3})	Bias^{2} (SD^{2})	r (SD)	Slope (SE)	RPIQ (SD)
PLSR	1.06 (0.010)	0.00 (0.042)	0.86 (0.004)	1.00 (0.010)	3.16 (0.11)
GAMM	1.10 (0.010)	0.04 (0.057)	0.84 (0.004)	0.99 (0.010)	3.04 (0.11)
NN	1.02 (0.010)	0.00 (0.040)	0.87 (0.004)	0.99 (0.010)	3.27 (0.10)

[1] RMSEV = root mean square error in validation data set; r = Pearson correlation between the actual observed value of the trait and predicted value of the trait; SD = standard deviation; SE = standard error.

[2] Values presented are values *1,000

[3] BCS units

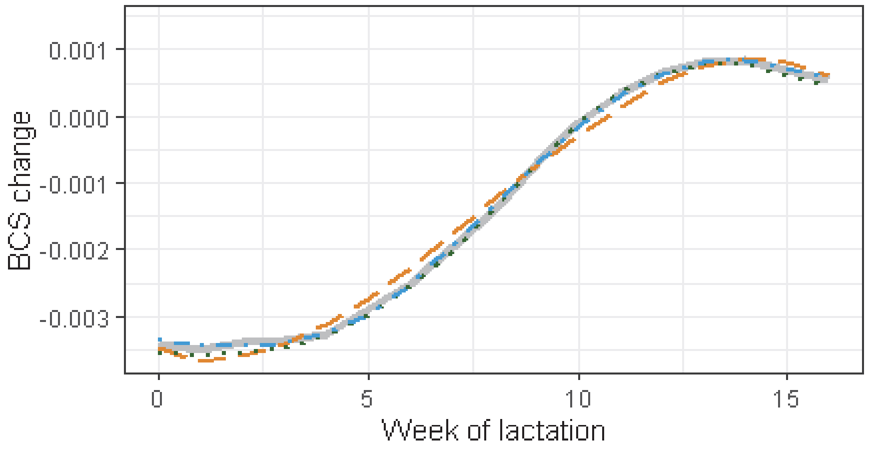


Figure 1. Body condition score (BCS) change (grey continuous line), ÄBCS predicted from partial least squares regression (orange long dashed line), generalized additive mixed model (green dotted line), and neural network (blue dot-dashed line).

Discussion

Body condition score, as well as ∆BCS, have been related to both fertility and health traits in lactating dairy cows (Berry *et al*., 2007; Buckley *et al*., 2003; Roche *et al*., 2007). Body condition score change is used by producers as an indicator of cow energy status, and, as demonstrated in the present study, it can be predicted from the MIR milk spectra.

One of the main challenge in predicting ∆BCS is accounting for scorer subjectivity. In fact, BCS evaluation, even if executed by trained scorers, is a subjective measure. Roche *et al*. (2004) studied the variability of international scoring systems, and reported a strong correlation between the scoring systems across countries. Moreover, in the same study, Roche *et al*. (2004) reported that the accuracy of the BCS assessed depended also on whether the assessment was just visual or a combination of visual plus tactile appraisal of the cow. This suggests that BCS assessment is somewhat scorer-dependent and may vary according to the scale and the method (visual or tactile appraisal) used by the scorer. In the present study, BCS was assessed by trained scorers using both visual and tactile appraisal. Using ∆BCS as opposed to BCS in the development of the prediction should negate the impact of mean score differences among scorers. As a further strategy to minimize the impact of the BCS uncertainty on the development of prediction methods, only ∆BCS predicted from the splines on the same day of an actual BCS record were retained. Nonetheless, considering the uncertainty of the trait being predicted, perfect or almost perfect prediction performance should not be expected.

McParland *et al*. (2014) previously predicted ∆BCS across 305 day of lactation in 1,018 lactating dairy cows from milk MIR using PLSR; they reported a correlation between the actual and the predicted ∆BCS of 0.75. Machine learning approaches have sometimes been shown to slightly improve the accuracy of predicting traits from milk MIR when compared to predictions developed using PLSR (Frizzarin *et al*., 2021b; Soyeurt *et al*., 2020). Partial least squares regression has the potential to be an interpretable statistical method, but assumes linearity between the trait and the latent variables of the spectra (Wold *et al*., 2001). Generalized additive mixed model had low variability in the ∆BCS predictions, and NN is less generalizable than simpler models, requires more tuning parameters identification, and requires large data sets. Therefore, it is important to consider the advantages and disadvantages of the different statistical methods when deciding which method to use for the final spectra predictions.

While the present study related individual cow milk MIR to ∆BCS, the prediction models could also possibly extend to the routinely taken herd bulk tank samples to assess mean herd energy status. Nevertheless, herd averages could hide individual cows mobilizing considerable body condition. Whereas, the present study was based on milk MIR samples taken weekly, this is usually not always practical in commercial farms. Nonetheless, technologies are being developed to routinely assess milk samples using different regions of the spectrum. These wavelengths are simply overtones of the mid-infrared region; hence, it could therefore possible to re-derive prediction equations for ∆BCS using data from different regions of the spectrum. Body condition score change can be used by the farmers to take specific animal decisions, like diet requirements, or whether inseminate a cow or not. Lastly, the prediction of ∆BCS after milk recording permits to generate a large amount of phenotypes which can be included in breeding programs (e.g., as correlated traits in a multi-trait genetic evaluation).

Conclusion

The results from the present study demonstrated the potential of using MIR spectra to predict ∆BCS in early lactating dairy cows. Body condition score change was predicted with a correlation between the actual and the predicted ∆BCS of 0.87 using NN and spectra obtained during morning milking. The prediction of ∆BCS from MIR spectra can be useful to obtain information on individual cow energy status, as well as routinely generate a phenotype for using in genetic evaluations.

Acknowledgments

This research was funded by a Science Foundation Ireland, Starting Investigator Research Grant, Infrared spectroscopy analysis of milk as a low cost solution to identify efficient and profitable dairy cows, 18/SIRG/5562, and has been supported in part by a research grant from Science Foundation Ireland and the Department of Agriculture, Food and Marine on behalf of the Government of Ireland under the Grant 16/RC/3835 (VistaMilk).

References

Bauman, D. E., and W. B. Currie. 1980. Partitioning of nutrients during pregnancy and lactation: A review of mechanisms involving homeostasis and homeorhesis. J. Dairy Sci. 63:1514–1529.

Berry, D. P., J. M. Lee, K. A. Macdonald, K. Stafford, L. Matthews, and J. R. Roche. 2007. Associations between body condition score, body weight, somatic cell count, and clinical mastitis in seasonally calving dairy cattle. J. Dairy Sci. 90:637–648.

Buckley, F., K. O'sullivan, J.F. Mee, R.D. Evans, and P. Dillon. 2003. Relationships among milk yield, body condition, cow weight, and reproduction in spring-calved Holstein-Friesians. J. Dairy Sci. 86(7), 2308-2319.

Butler, W. R., and R. D. Smith. 1989. Interrelationships between energy balance and postpartum reproductive function in dairy cattle. J. Dairy Sci. 72:767-783.

De Marchi M.D., Toffanin V., Cassandro M. and Penasa M. 2014. Invited review: mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97, 1–16.

Edmonson, A. J., I. J. Lean, L. D. Weaver, T. Farver, and G. Webster. 1989. A body condition scoring chart for Holstein dairy cows. J. Dairy Sci. 72:68–78.

Frizzarin, M., O'Callaghan, T.F., Murphy, T.B., Hennessy, D. and Casa, A., 2021a. Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets. J. Dairy Sci., 104(12), pp.12394-12402.

Frizzarin M., I.C. Gormley, D.P. Berry, T.B. Murphy, A. Casa, A. Lynch, and S. McParland, 2021b. Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. J. Dairy Sci. 2021, 104, 7438–7447.

Grelet, C., J.F. Pierna, P. Dardenne, V. Baeten, and F. Dehareng, 2015. Standardization of milk mid-infrared spectra from a European dairy network. J. Dairy Sci, 98(4), pp.2150-2160.

Grelet, C., J.F. Pierna, P. Dardenne, H. Soyeurt, A. Vanlierde, F. Colinet, C. Bastin, N. Gengler, V. Baeten, and F. Dehareng, 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J. Dairy Sci., 100(10), pp.7910-7921.

McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. Pryce, and D.P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. J. Dairy Sci 97(9):5863-5871.

Mevik B.H, R. Wehrens and K. H. Liland. 2019. pls: Partial Least Squares and Principal Component Regression. R package version 2.7-2.

Nogalski, Z., Wrotski, M., Sobczuk-Szul, M., Mochol, M., and Pogorzelska, P. 2012. The Effect of Body Energy Reserve Mobilization on the Fatty Acid Profile of Milk in High-yielding Cows. Asian-Australasian Journal of Animal Sciences. Asian Australasian Association of Animal Production Societies.

Perez Rodriguez P. and D. Gianola. 2020. brnn: Bayesian Regularization for Feed-Forward Neural Networks. R package version 0.8.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Roche, J. R., P. G. Dillon, C. R. Stockdale, L. H. Baumgard, and M.J. VanBaale. 2004. Relationships among international body condition scoring systems. J. Dairy Sci. 87:3076–3079

Roche, J. R., K. A. Macdonald, C. R. Burke, J. M. Lee, and D. P. Berry. 2007. Associations among body condition score, body weight, and reproductive performance in seasonal-calving dairy cattle. J. Dairy Sci. 90:376–391.

Roche, J. R., N. C. Friggens, J. K. Kay, M. W. Fisher, K. J. Stafford, and D. P. Berry. 2009. Invited review: body condition score and its association with dairy cow productivity, health, and welfare. J. Dairy Sci. 92:5769–5801.

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. J. Dairy Sci 89(9):3690-3695.

Soyeurt H., C. Grelet, S. McParland, M. Calmels, M. Coffey, A. Tedde, P. Delhez, F. Dehareng, and N. Gengler. A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra. J. Dairy Sci 103.12 (2020): 11585-11596.

Wold, S., M. Sjostrom, and L. Eriksson. 2001. PLS-Regression: A basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58:109–130.

Wood S. and F. Scheipl. 2020. gamm4: Generalized Additive Mixed Models using 'mgcv' and 'lme4'. R package version 0.2-6.

Valorization of milk spectra: data mining of milk infrared spectra to assess transition success

M. Bahadi[1,2], D. Warner[1], R. Moore[1], R. Lacroix[1], R. Cue[2] and D.E. Santschi[1]

[1]Lactanet, 555 Boulevard des Anciens-Combattants, H9X 3R4,
Sainte Anne de Bellevue, QC, Canada
Corresponding Author: mbahadi@lactanet.ca
[2]McGill University, Department of Animal Science, 21111 Lakeshore, H9X 3V9,
Sainte Anne de Bellevue, QC, Canada

Abstract

The transition period is recognized as a critical phase in a cow's lactation. Poor transition can often be associated with clinical or subclinical diseases, but there is still a proportion of cows where poor transition cannot be linked to routinely measurable metabolites in milk. Also, millions of milk infrared spectra are generated in milk recording laboratories every year. These spectra contain comprehensive information about milk chemical composition. The objective of this study was to evaluate whether milk spectra can be mined in search for minor milk components that are not routinely measured in monthly milk samples and that might be potential biomarkers to assess transition success. First test day records within lactation and their corresponding spectra between 2015 and 2020 were extracted from Lactanet's database for Holstein cows in Québec, Canada. A categorical variable was created as a proxy for transition management based on the value of Transition Cow Index (TCI). ANOVA–simultaneous component analysis+ (ASCA+) was used to test the effect of the TCI category on spectral variability. In the first round of analysis, spectra of samples collected during the first two weeks of lactation (N=238,773) had the highest variability attributed to the TCI category, which peaked on 8-11 DIM at 3.02%. This variability falls to 1.41% on week 4 of lactation. Spectral variabilities attributed to the other studied factors, namely DIM and parity (2, 3+), were <1% and <0.5%, respectively. The second round of analysis included 41,464 spectra of samples collected during the first two weeks of lactation. The results of the analysis revealed that low TCI category had direct correlation with spectral features of milk fat and protein and inverse correlation with those of lactose. In addition, direct relationship was observed between spectral features that can be attributed to milk fat, protein, creatine, phosphate, sulfur containing compounds and trans fatty acids.

These findings suggest that more frequent milk sampling is needed during the first two weeks of lactation and monitoring additional minor milk components might be useful in assessing transition success. To conclude, milk infrared spectra represent a rich source of information regarding the chemical composition of milk, and they can be mined to gain insights and detect trends to assess transition success in dairy cows.

Keywords: Cow transition, FTIR spectroscopy, biomarkers.

Introduction

A good transition cow management is important as it affects the entire productive life of a dairy cow. The Transition Cow Index® (TCI) was developed by Ken Nordlund at the University of Wisconsin (Nordlund, 2006). The principle involves a 13-factor equation used to predict how much milk a cow is expected to produce during her 305-day lactation. Factors included in the equation are environmental (calving month), related to the cow (breed, parity) or related to her current (days in milk at first test, lactation start reason) or previous lactation (production, number of days dry, lactation start reason, lactation duration, last test day somatic cell score). At the cow's first test date, the actual projected 305-day milk yield is compared to this prediction. If the cow's production is exactly as predicted, the TCI is 0. If the cow produces more than predicted, the TCI will be positive; if less, then the TCI will be negative. A TCI value can be calculated for all second and greater lactation cows that are between 5 and 39 DIM at their first test day. The TCI is an objective way of evaluating the start of lactation for each cow. Since a good start to lactation is generally a promising sign for a good performance in overall lactation, an increase in the TCI suggests that cows will have a higher milk yield over the coming lactation. Generally speaking, a 100-point increase in a herd's average TCI corresponds to a milk yield increase of 93 kg of milk per cow for 305 days of production (Lactanet data, 2014).

Previous work has demonstrated that a very high proportion of 1st lactation cows have elevated milk beta-hydroxybutyrate (BHB) on their first test (Santschi et al., 2016a) indicating an excessive body reserve mobilization which could be linked to poor transition, excessive stress related to the onset of lactation, or a normal biological response. The ideal detection method for subclinical disorders in early lactation cows would be blood monitoring of specific metabolites. However, this technique is laborious, expensive, and intrusive. Recently, several studies suggested the use of milk spectral information and Fourier-transformed infrared (FTIR) spectroscopy as a metabolic profiling tool, specifically in early lactation (de Roos et al., 2007, Santschi et al., 2016a, Pryce et al., 2019). Around the world, routine testing for BHB has become a fast and inexpensive way to monitor herd-level prevalence of hyperketonaemia (Santschi et al., 2016b). It is hypothesized that routine testing of early lactation milk could therefore become an effective and affordable screening tool to identify individual cows needing specific attention. Recent studies have tried to predict specific metabolites with variable precision (Barbano et al., 2015, Pape et al., 2018, Luke et al., 2019), often by using FTIR to predict a single blood component. The present study aims at looking at overall spectral signature rather than trying to predict specific metabolites. This approach aims at screening the differences in spectral signature, and then eventually identify the possible metabolites involved, using a historical database of several hundred thousand Canadian Holstein cows, and using TCI as the classification variable to group cows based on their status.

Materials and methods

Assembly of the dataset

First test within lactation records and their corresponding spectra between 2015 and 2020 were extracted from Lactanet's database. The total number of extracted records was 238,773. A categorical factor was created, TCI_CAT, and three levels were defined for it as follow: 1) high, which contained records with TCI value $\geq$66th percentile and they represented successful transition, 2) mid, which contained records with TCI value >33rd percentile and <66th percentile and they were dropped to enhance the contrast in differences between records in the other two categories, 3) low, which contained records with TCI value $\leq$33rd percentile and they represented poor transition.

Testing the effect of the TCI category on spectral data

ANOVA simultaneous component analysis + (ASCA+) was chosen to test the effect of the TCI category on the spectral dataset. In the first round of analysis, the spectral dataset was divided into subsets according to days in milk (DIM): 5-7 DIM, 8-11 DIM, 12-14 DIM, week 3, week 4, week 5&6. The ASCA+ model contained the following factors: the TCI category, parity groups (second, third and higher) and DIM. In the second round of analysis, data and spectral records were limited to those that were collected during the first two weeks of lactation (N=41,464) and the spectral dataset was divided into subsets according to parity (i.e., all, second, third and higher).

Attributing milk components to the TCI category effect

Inspection of the loading spectrum of the TCI category factor allowed the association of specific milk components with the TCI category effect. High loadings, positive or negative, at characteristic wavenumbers of milk components were observed and were interpreted according to methodology described elsewhere (Bahadi et al., 2021).

Results and discussion

The first round of analysis revealed that milk spectra could capture more variability related to cow transition in the first two weeks of lactation (Table 1), which peaked in 8-11 DIM at 3.02%. This variability falls to 1.41% on week 4 of lactation. Spectral variabilities attributed to the other studied factors, namely DIM and parity (i.e., second, third and higher), were <1% and <0.5%, respectively. This observation means that the current first test date window of 6 weeks might not be tight enough to capture differences in milk composition between cows that do well or poorly during the transition period. To better understand these differences, the first test date window might need to be shortened to less than 6 weeks, preferably to two weeks of lactation, in order to better capture these differences in milk composition.

Inspection of the loading spectrum of the TCI_CAT variable reveals positive loadings at characteristic wavenumbers of milk fat and protein, and negative loadings at wavenumbers attributed to lactose. This observation suggests the existence of an inverse relationship between fat and protein, on one hand, and lactose, on the other hand, in milk samples belonging to either low or high TCI category. Indeed, when we inspected the means of fat, protein and lactose that were determined by the lab for the studied samples, we observed that milk samples from cows classified in the low TCI category had higher fat and protein content and lower lactose content in comparison with milk samples coming from cows in the high TCI category. In other words, TCI values had an inverse relationship with fat and protein content and a direct one with lactose content in milk samples.

In the second round of analysis, variability in minor milk components' spectral features was observed and it was attributed to differences in milk composition in samples coming from cows assigned to low and high TCI categories. These minor milk components are sulfur containing compounds, trans fatty acids, creatine, and phosphate. Hence, these components might be candidate biomarkers for monitoring transition success. In fact, sulfur containing compounds were already reported as probable biomarkers to monitor transition success in studies done on cows' blood plasma and liver (Zhou et al., 2017). In addition, trans fatty acids have already been reported in the literature as contributors to inflammation in bovine mammary epithelial cells (Rezamand and McGuire, 2011). The variability in lactose, sulfur containing compounds and trans fatty acids was highest during the first two weeks of lactation and in milk samples coming from cows in their third parity or higher. Cows that suffer from poor transition in a lactation become more susceptible to inflammation and health issues in subsequent

Table 1. Percentage of variance in the spectral data described by several factors of the ASCA+ model.

Term	5-7 D[1]	8-11 D	12-14 D	W3[2]	W4	W5&6
Parity	0.48	0.34	0.44	0.25	0.31	0.28
DIM	0.80	0.98	0.16	0.80	0.25	0.24
TCI_CAT	2.10	3.02	2.14	1.65	1.41	1.88

[1] D = Days
[2] W = Week

lactations, which might mean more variability in candidate transition biomarkers in cows with higher parities.

Conclusion

The findings of this study suggest that the first test date window might need to be shortened to less than 6 weeks and it is preferable to do the first test during the first two weeks of lactation in order to capture the differences in milk composition between cows that do well or poorly during their transition period. In addition, monitoring more minor milk components might be useful in assessing transition success. Milk infrared spectra represent a rich source of information regarding the chemical composition of milk, and they can be mined to gain insights and detect trends to assess transition success in dairy cows.

Acknowledgement

This work was supported by Mitacs through the Mitacs Accelerate program.

References

Bahadi, M., A. A. Ismail, and E. Vasseur. 2021. Fourier Transform Infrared Spectroscopy as a Tool to Study Milk Composition Changes in Dairy Cows Attributed to Housing Modifications to Improve Animal Welfare. Foods 10(2):450.

Barbano, D., P. Cree, T. Overton, H. Dann, and R. Grant. 2015. Prediction of blood nonesterified fatty acid (NEFA) by FTIR analysis of individual cow milk samples. Journal of Dairy Science 98(1):553.

de Roos, A. P. W., H. J. C. M. van den Bijgaart, J. Hørlyk, and G. de Jong. 2007. Screening for Subclinical Ketosis in Dairy Cattle by Fourier Transform Infrared Spectrometry. Journal of Dairy Science 90(4):1761-1766.

Luke, T., S. Rochfort, W. Wales, V. Bonfatti, L. Marett, and J. Pryce. 2019. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. Journal of Dairy science 102(2):1747-1760.

Nordlund, K. 2006. Transition cow index™. Pages 139-143 in Proc. American Association of Bovine Practitioners Proceedings of the Annual Conference.

Pape, A., H. Dann, D. Barbano, and R. Grant. 2018. Detection of health problems by changes in milk estimated blood nonesterified fatty acids (NEFA) and milk fat, protein, and fatty acids. Journal of Dairy Science 101:39.

Pryce, J., T. Nguyen, P. Ho, T. Luke, S. Rochfort, W. Wales, P. Moate, L. Marett, G. Nieuwhof, and M. Abdelsayed. 2019. Applying next generation phenotyping strategies for genetic gain in dairy cattle. Pages 390-393 in Proc. Proc. Assoc. Advmt. Anim. Breed. Genet.

Rezamand, P. and M. McGuire. 2011. Effects of trans fatty acids on markers of inflammation in bovine mammary epithelial cells. Journal of Dairy science 94(1):316-320.

Santschi, D., R. Lacroix, J. Durocher, M. Duplessis, R. Moore, and D. Lefebvre. 2016a. Prevalence of elevated milk β-hydroxybutyrate concentrations in Holstein cows measured by Fourier-transform infrared analysis in Dairy Herd Improvement milk samples and association with milk yield and components. Journal of Dairy science 99(11):9263-9270.

Santschi, D., R. Lacroix, R. Moore, F. Miglior, and D. Lefebvre. 2016b. Impacts of early lactation hyperketonemia on reproduction and 305-d milk production. Journal of Animal Science 94:611-611.

Zhou, Z., E. Trevisi, D. Luchini, and J. Loor. 2017. Differences in liver functionality indexes in peripartal dairy cows fed rumen-protected methionine or choline are associated with performance, oxidative stress status, and plasma amino acid profiles. Journal of Dairy science 100(8):6720-6732.

Within day variation in milk and blood metrics for hyperketonemic and non-hypoketonemic dairy cows

C. Seely[1], J.A.A. McArt[1] and D.M. Barbano[2]

[1]Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, NY 14853, USA
[2] Northeast Dairy Foods Research Center, Department of Food Science, Cornell University, Ithaca, NY 14853, USA
Corresponding Author: Barbano1@aol.com

Abstract

Dairy cows often enter a state of energy deficit in early lactation, leading to an increase in plasma concentrations of non-esterified fatty acids (NEFA) and beta-hydroxybutyrate (BHB). Currently, diagnosis of excessive energy deficit is done on farms using handheld blood BHB meters. However, this process is laborious and can become costly when used as a whole-herd screening method. Several studies have investigated the use of Fourier transform mid-infrared (FTIR) estimates to predict excessive energy deficit, but these studies relied on a single, test-day DHIA milk sample with no knowledge of actual blood NEFA or BHB concentrations. We determined the diurnal variation in plasma NEFA and BHB as well as FTIR estimates of milk BHB, milk predicted blood NEFA, and milk fatty acids with particular focus on differences between groups of cows that were hyperketonemic or non-hyperketonemic. We collected blood samples every 2 h for 5 consecutive days from 28 multiparous Holstein cows that were between 3 and 9 days in milk. Cows were housed in a tie-stall facility and offered free choice access to water and a TMR that was delivered once a day at 0900 h. Blood samples were analyzed for BHB and NEFA concentrations, and cows were classified into hyperketonemia groups based on their average daily BHB concentration. If a cow's average daily BHB was $\geq$1.2 mmol/L for $\geq$3 study days, she was assigned to the hyperketonemia group (n=13). Alternatively, if her average daily BHB was $\geq$1.2 mmol/L for $\leq$2 study days, she was assigned to the non-hyperketonemia group (n=15). We found clear and consistent diurnal patterns in plasma BHB and NEFA as well as FTIR estimates of milk BHB, milk predicted blood NEFA, and milk fatty acids. Interestingly, these diurnal differences were much more predictable when analyzing milk, with a greater ability to separate hyperketonemic from non-hyperketonemic cows. Our results support the use of FTIR estimates of milk BHB and milk predicted blood NEFA as a tool in diagnosing HYK, however time relative to feeding should be considered when analyzing results. Milk fatty acid metrics on a relative basis may also be useful to separate hyperketonemic from non-hyperketonemic cows. In particular, these results support the use of high frequency milk monitoring and measurement to detect alterations in early lactation health of dairy cows.

Introduction

Dairy cows often enter a state of energy deficit in early lactation, leading to an increase in plasma concentrations of non-esterified fatty acids (NEFA) and beta-hydroxybutyrate (BHB). Currently, diagnosis of excessive energy deficit is done on farms using handheld blood BHB meters. However, this process is laborious and can become costly when used as a whole-herd screening method. Several studies have investigated the use of Fourier transform mid-infrared (FTIR) estimates to predict excessive energy deficit through milk (Denis-Robichaud *et al*., 2014; Santschi *et al*., 2016; Bach *et al*., 2019), but many of these studies relied on a single, test-day DHIA milk sample with no knowledge of actual blood NEFA or BHB concentrations. Here we present our investigation of the diurnal variation in plasma NEFA and BHB as well as FTIR estimates of milk BHB and predicted blood NEFA, with particular interest in differences between groups of cows that were hyperketonemic or non-hyperketonemic. This information will improve knowledge and usability of on-farm testing results and promote discussion of the benefits of routine milk testing and analysis.

Study design and results

We collected blood samples every 2 h for 5 consecutive days from 28 multiparous Holstein cows that were between 3 and 9 days in milk. Cows were housed in a tie-stall facility and offered free choice access to water and a TMR that was delivered once a day at 0900 h. Blood samples were analyzed for BHB and NEFA concentrations, and cows were classified into hyperketonemia groups based on their average daily BHB concentration. If a cow's average daily BHB was $\geq$1.2 mmol/L for $\geq$3 study days, she was assigned to the hyperketonemia group (n=13). Alternatively, if her average daily BHB was $\geq$1.2 mmol/L for $\leq$2 study days, she was assigned to the non-hyperketonemia group (n=15).

Blood results

We saw the lowest concentrations of BHB just prior to feeding, at 0700 h, with a steady rise following feed delivery (Figure 1A). Not surprisingly, BHB was higher in the hyperketonemic cows than the non-hyperketonemic cows (Figure 1B). Unlike BHB however, we saw a peak in NEFA just prior to feeding at 0700 h, with concentrations falling quickly after feed delivery (Figure 1C). The hyperketonemic cows had greater concentrations of NEFA than the non-hyperketonemic cows (Figure 1D).

To understand the effect of hyperketonemic on the daily fluctuations of BHB and NEFA, we calculated the difference between the daily maximum and minimum concentrations for each metabolite by hyperketonemia group. The hyperketonemic cows experienced a nearly two-fold greater difference between daily maximum and minimum BHB concentration as compared to the non-hyperketonemic cows. Interestingly, the difference between daily maximum and minimum concentrations of NEFA were relatively similar for both the hyperketonemic and non-hyperketonemic cows.

Milk results

We saw similar diurnal findings with mid-FTIR milk predicted metabolites, however with a general lag in peak or nadir concentrations than blood. The lowest milk BHB and milk predicted blood NEFA concentrations were at the morning milking just prior to feeding (Figure 2A, 2C). As for blood, predicted milk BHB and milk predicted blood NEFA were higher in hyperketonemic than non-hyperketonemic cows (Figure 2B, 2D). However, unlike blood, difference in milk BHB between hyperketonemic groups was more consistent, and the pattern of diurnal variation in milk predicted blood NEFA never overlapped between the two groups.

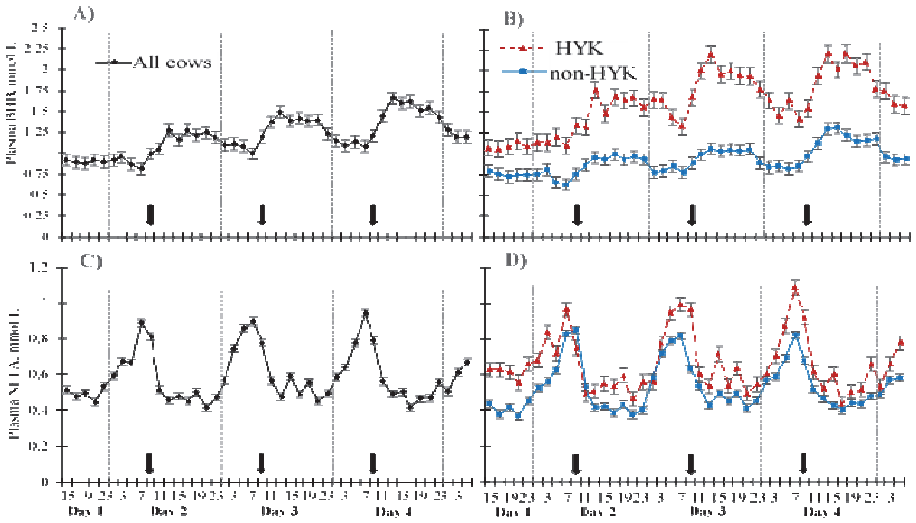


Figure 1. Concentrations of BHB and NEFA from multiparous Holstein cows classified as hyperketonemic (HYK; average daily mean BHB ≥1.2 mmol/L for ≥3 study days, red dashed line) or non-hyperketonemic (non-HYK; average daily mean BHB ≥1.2 mmol/L for ≤2 study days, solid blue line). Black arrows indicate time of feed delivery (Adapted from Seely et al., Journal Dairy Science, In Press).

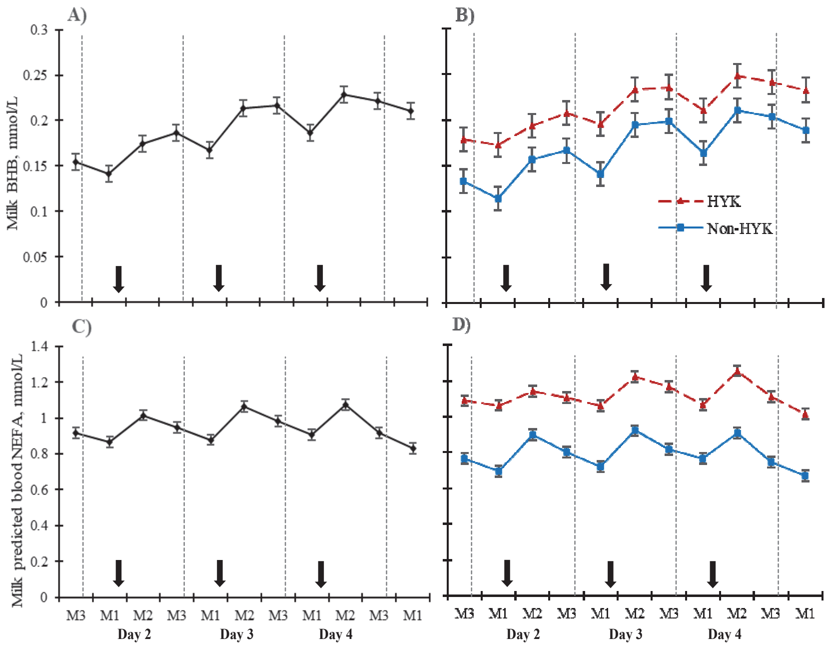


Figure 2. Concentrations of mid-FTIR milk predicted metabolites (milk BHB and milk predicted blood NEFA) from multiparous Holstein cows classified as hyperketonemic (HYK; average daily mean BHB ≥1.2 mmol/L for ≥3 study days, red dashed line) or non-hyperketonemic (non-HYK; average daily mean BHB ≥1.2 mmol/L for ≤2 study days, solid blue line). Black arrows indicate time of feed delivery; M1 = morning milking, M2 = afternoon milking, M3 = evening milking (Unpublished data from Seely, McArt, and Barbano).

Cows were milked 3 times per day. Therefore, each milk sample theoretically represents the average of what happened in the blood for the 8 h period prior to milking. We averaged the 2 h blood testing data over each 8 h period, prior to each milking, to achieve better correspondence of the time period for milk and blood results. Milk predicted blood NEFA, milk BHB, and milk fatty acids were measured at each milking using a mid-FTIR milk analysis (Delta FTA, Perkin-Elmer Corp., Drachten, The Netherlands). The data for milk predicted blood NEFA and blood NEFA over consecutive milkings for the hyper and non-hyperketonemic groups of cows are shown in Figures 3 and 4. Both the blood NEFA and milk estimated blood NEFA cycled during each 24 h period with a slight lag in timing of the cycling. Both the blood NEFA and the milk estimated blood NEFA clearly separated the two groups of cows.

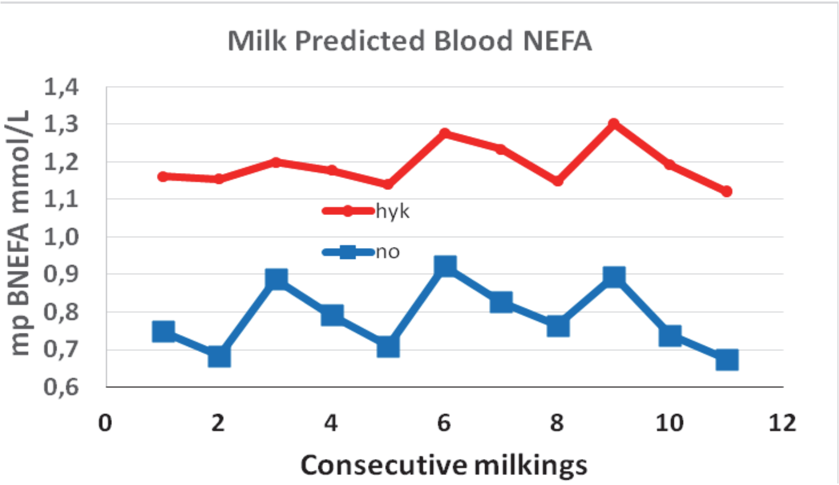


Figure 3. Milk estimated blood NEFA over consecutive milkings.

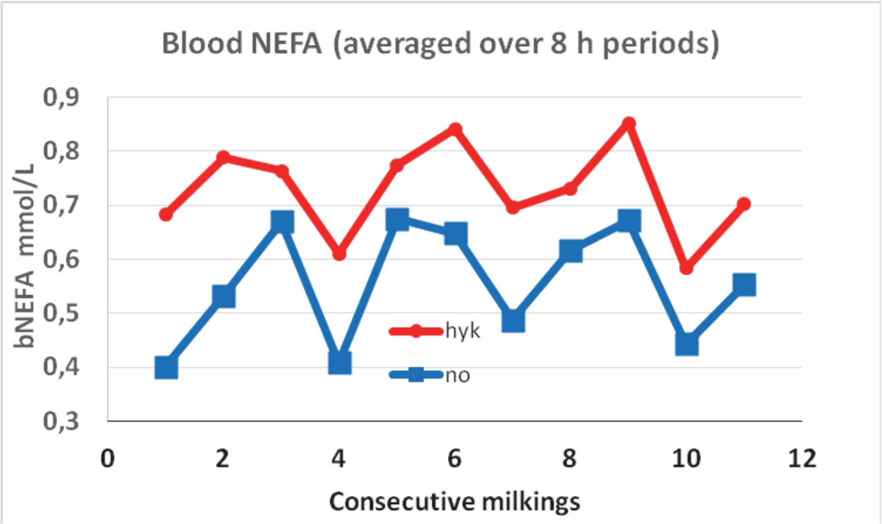


Figure 4. Blood NEFA (2 h samples averaged across the 8 h period prior to milking) plotted as a function of consecutive milkings.

The concentration of milk BHB was also measured by infrared milk analysis and concentration in milk also cycled (data not shown).

Milk fatty acids (denovo, mixed origin, and preformed) were also measured at every milking by mid-infrared milk analysis (Figures 5, 6, and 7), as described by Wojciechowski *et al.,* 2016 and Woolpert *et al.,* 2016. The comparison of relative concentration for the 3 different groups of milk fatty acids between hyper and non-hyperketonemic groups of cows was clearly separated for all three milk fatty acid metrics. The non-hyperketonemic cows had higher relative concentrations of de novo and mixed origin milk fatty acids and lower performed milk fatty acids than the

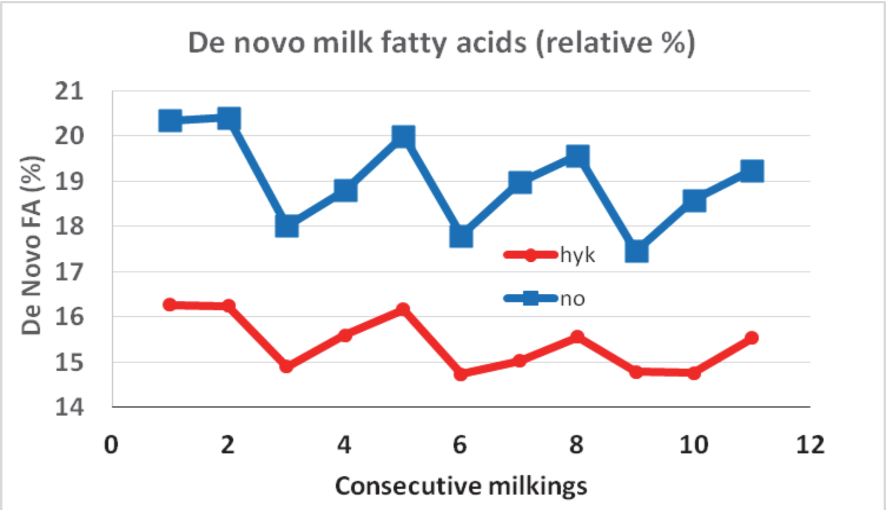


Figure 5. De novo milk fatty acids (relative %) measured by mid-infrared milk analysis. Fresh feed offered immediately after milking 3, 6, and 9.

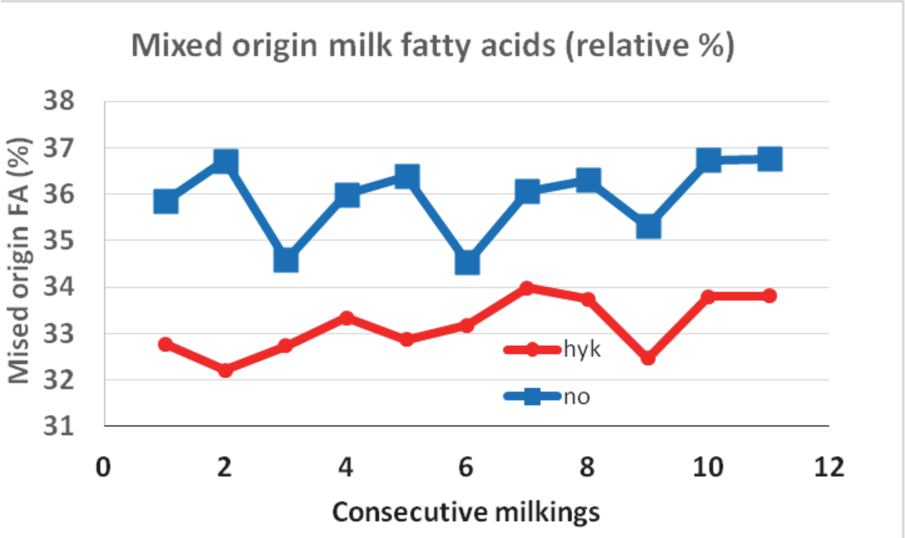


Figure 6. Mixed origin milk fatty acids (relative %) measured by mid-infrared milk analysis. Fresh feed offered immediately after milking 3, 6, and 9.

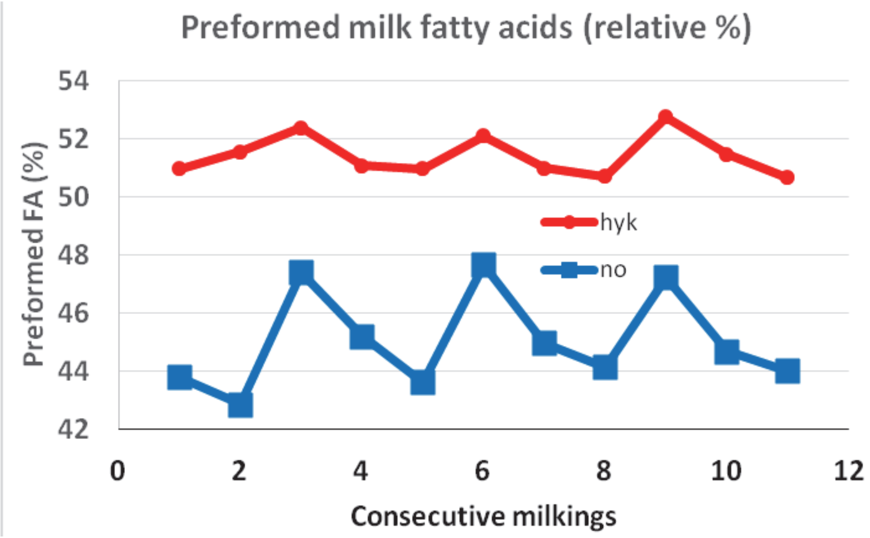


Figure 7. Preformed milk fatty acids (relative %) measured by mid-infrared milk analysis. Fresh feed offered immediately after milking 3, 6, and 9.

hyperketonemic cows. The cycle phasing of the relative concentrations of the de novo and mixed origin fatty acids had the opposite phasing when compared with the phasing of the preformed milk fatty acid cycling. Cycling of the relative concentration of milk fatty acid groups (Figures 5, 6, 7) was related to cycling of blood NEFA and the cycling of the fatty acid groups was consistent with the milk estimated blood NEFA cycling (Figure 3). When milk estimated blood NEFA was at a maximum of a cycle, preformed fatty acids were also at the maximum. Immediately before the cows were given fresh feed, milk estimated blood NEFA and relative concentration of milk preformed fatty acids were at a maximum and the de novo and mixed origin fatty acids were at a minimum.

We hypothesize that the differences between peak and nadir blood and milk metabolites are due to milk having a higher correlation with an 8-hour average of blood metabolite concentrations rather than a single blood sample. This makes biological sense and also supports the idea that milk analysis might be an improved method of representing a cow’s overall energy status than a single snapshot in time as currently provided with blood sampling.

Acknowledgement

We would like to acknowledge Dr. Kathryn Bach, the post-doctoral associate who assisted with study organization. Our study was funded by the USDA National Institute of Food and Agriculture, Hatch project no. 1017096.

References

Bach, K. D., D. M. Barbano, and J. A. A. McArt. 2019. Association of mid-infrared-predicted milk and blood constituents with early-lactation disease, removal, and production outcomes in Holstein cows. J. Dairy Sci. 102:10129-10139.

Denis-Robichaud, J., J. Dubuc, D. Lefebvre, and L. DesCoteaux. 2014. Accuracy of milk ketone bodies from flow-injection analysis for the diagnosis of hyperketonemia in dairy cows. J. Dairy Sci. 97:3364-3370.

Santschi, D.E.,R. Lacroix, J. Durocher, M. Duplessis, R.K. Moore, and D.M. Lefebvre. 2016. Prevalence of elevated milk beta-hydroxybutyrate concentrations in Holstein cows measured by Fourier-transform infrared analysis in Dairy Herd Improvement milk samples and association with milk yield and components. J. Dairy Sci. 99:9263–9270.

Wojciechowski, K.L, and D.M. Barbano. 2016. Prediction of fatty acid chain length and unsaturation of milk fat by mid-infrared milk analysis. J. Dairy Sci. 99:8561-8570.

Woolpert, M.E., H. M. Dann, K.W. Cotanch, C. Melilli, L.E. Chase, R.J. Grant, and D.M. Barbano. 2016. Management, nutrition, and lactation performance are related to bulk tank milk de novo fatty acid concentration on northeastern US dairy farms. J. Dairy Sci. 99:8486-8497.

New traits predicted from milk mid-infrared spectra to reduce incidence of subclinical ketosis

A. Köck[1], K. Schodl[1 2], B. Fuerst-Waltl[2], H. Schwarzenbacher[1], L.M. Dale[3], A. Werner[3], M. Mayerhofer[1], F.J. Auer[4], C. Grelet[5], J. Sölkner[2], L. Rienesl[2], N. Gengler[6], J. Leblois[7] and C. Egger-Danner[1]

[1]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89, 1200 Vienna, Austria
Corresponding Author: koeck@zuchtdata.at
[2]University of Natural Resources and Life Sciences, Vienna, 1180 Vienna, Austria
[3]Regional State Association for Performance and Quality Inspection in Animal Breeding of Baden Württemberg, Heinrich-Baumann-Str. 1-3, 70190 Stuttgart, Germany
[4]LKV-Austria, Dresdner Str. 89, 1200 Vienna, Austria
[5]Walloon Agricultural Research Centre, Gembloux, Belgium
[6]University of Liege, Gembloux Agro-Bio Tech, Gembloux, Belgium
[7]EEIG European Milk Recording (EMR), Ciney, Belgium

Abstract

Ketosis is the most frequent metabolic disease in dairy cows. Recently, several new mid-infrared (MIR) traits have been derived that can be predicted from routine milk samples and provide a more accurate indication of subclinical ketosis than the commonly used fat-to-protein ratio, such as KetoMIR and other MIR-predicted traits (e.g. blood beta-hydroxybutyrate, acetone, citrate). KetoMIR was developed by LKV Baden-Württemberg based on ketosis diagnoses. KetoMIR is a three-class ketosis index: 1 = low ketosis risk, 2 = moderate ketosis risk, and 3 = high ketosis risk. The increased ketosis risk based on the KetoMIR index was associated with lower average herd milk yield (-1,975 kg milk). The interval from calving to first service was prolonged by +36 days, as was the calving interval with +58 days. Mean herd somatic cell count in first and higher lactations was increased by 60,000 and 134,000 cells/ml, respectively. So far, KetoMIR results have only been used for herd management. Feeding advisors use this new MIR trait to assess and, if necessary, adjust the feeding situation on the farm in the dry cow period and early lactation. Furthermore, a MIR equation for beta-hydroxybutyrate in blood was derived, which has already been validated on 49 Austrian farms and 670 dairy cows. For this purpose, capillary blood was analyzed for beta-hydroxybutyrate concentration in all cows during milk recording in early lactation (1st and 2nd test day after calving) using a handheld device (WellionVet BELUA, MED TRUST Handels GmbH, Marz, Austria). The result from the handheld device was considered as the gold standard for detecting subclinical ketosis (betahydroxybutyrate concentration > 1.2 mmol/l). Blood beta-hydroxybutyrate predicted from MIR had a sensitivity of 56% and a specificity of 81% for detecting cows with subclinical ketosis. Currently, data from Austria are being integrated into the MIR equation for beta-hydroxybutyrate to improve the equation. First genetic analyses showed high heritabilities between 0.16 and 0.30 for MIR-predicted traits. The moderate

to high genetic correlations between MIR-predicted traits and subclinical ketosis suggest that consideration of these traits in selection would help to reduce subclinical ketosis.

Keywords: mid-infrared, subclinical ketosis, herd management, genetic selection

Introduction

Metabolic disorders occur more frequently after calving. In this phase, the energy demand due to milk production is higher than the feed intake, resulting in a more or less pronounced negative energy balance. Although the majority of metabolic disorders are not visible, but are present subliminally, these metabolic disturbances lower the cow's resistance and increase, for example, the risk of mastitis or hoof diseases (Pieper and Mahlkow-Nerge, 2017).

Ketosis is the most common metabolic disease of the dairy cow. Since affected cows usually show no signs of disease at all, detection is very difficult. Recently, several new mid-infrared (MIR) traits have been derived that can be predicted from routine milk samples and provide a more accurate indication of subclinical ketosis than the commonly used fat-to-protein ratio, such as KetoMIR and other MIR-predicted traits (e.g. beta-hydroxybutyrate (BHB) and acetone).

KetoMIR

KetoMIR was developed by the LKV Baden-Württemberg (Dale *et al*., 2018). Using veterinary ketosis diagnoses and the MIR spectra of milk samples from the first 120 days of lactation from milk recording, KetoMIR, a three-level ketosis index, was developed (Dale *et al*., 2018). Class 1 indicates low ketosis risk, class 2 indicates medium ketosis risk, and class 3 indicates high ketosis risk.

The KetoMIR index has a sensitivity of 68% and a specificity of 81% to detect clinical ketosis (Dale *et al*., 2018). KetoMIR is an alert system to help in herd management during the first 120 days of lactation, but it is not diagnostic.

In Austria, the average frequency of cows with a positive KetoMIR result is 14% at the farm level. Highly elevated frequencies of >30% are found in 8.6% of farms. The increased ketosis risk based on the KetoMIR index was associated with lower average herd milk yield (-1,975 kg milk). Mean herd somatic cell count in first and higher lactations was increased by 60,500 and 134,400 cells/ml, respectively. The interval from calving to first service was prolonged by +36.5 days, as was the calving interval with +58.2 days.

Blood beta-hydroxybutyrate

A MIR equation for beta-hydroxybutyrate (BHB) in blood was available from European Milk Recording (EMR), which has already been validated on 49 Austrian farms and 670 dairy cows. For this purpose, capillary blood was analyzed for BHB concentration in all cows during milk recording in early lactation (1st and 2nd test day after calving) using a handheld device (WellionVet BELUA, MED TRUST Handels GmbH, Marz, Austria). The result from the handheld device was considered as the gold standard for detecting subclinical ketosis (BHB > 1.2 mmol/l). Blood BHB predicted from MIR had a sensitivity of 55.9% and a specificity of 81.4% for detecting cows with subclinical ketosis, which was higher than for the commonly used fat-protein ratio > 1.5 (Table 1).

Genetic parameters

Heritabilities and genetic correlations for BHB measured with the handheld device and MIR predicted ketosis risk traits are shown in Table 2. Heritabilities for all traits were high, 0.26 for BHB results gained by the handheld device and between 0.16 and 0.30 for MIR predicted traits. The heritability estimates found in our study were consistent with previous literature results (Belay *et al*., 2017; Hamann *et al*., 2017; Benedet *et al*., 2020). The genetic correlations between BHB concentration from the handheld device and the traits predicted by MIR ranged from 0.60 to 0.73. Genetic correlations between BHB results gained by a handheld device and MIR predicted traits recorded at the same time could not be found in the literature. However genetic correlations with clinical ketosis from veterinary diagnoses were available. A moderate genetic correlation of 0.47 between predicted blood BHB at first test day and clinical ketosis was reported by Belay *et al*. (2017). Because the standard errors of the estimates in our study were high, the results should be taken with caution. However, the moderate to high genetic correlations between MIR-predicted traits and subclinical ketosis suggest that consideration of these traits would reduce subclinical ketosis.

Conclusion

Recently, several new MIR traits have been derived that can be predicted from routine milk samples and provide a more accurate indication of subclinical ketosis than the commonly used fat-to-protein ratio, such as KetoMIR, beta-hydroxybutyrate and acetone. Preliminary results suggest that these new traits could be used in both herd management and breeding programs to reduce the incidence of subclinical ketosis.

Table 1. Sensitivity and specificity given that cows with subclinical ketosis have a beta-hydroxybutyrate-concentration > 1.2 mmol/l using the handheld device.

	Sensitivity	Specificity
Blood beta-hydroxybutyrate > 200µmol/L	55.9	81.4
Fat-protein ratio > 1.5	39.2	83.1

Table 2. Heritabilities (on diagonal) and genetic correlations (above diagonal) with standard errors in brackets.

	BHB – handheld device	KetoMIR	BHB-MIR	Aceton-MIR	Fat-protein-ratio
BHB – handheld device	0.26	0.61 (0.35)	0.60 (0.32)	0.73 (0.33)	0.60 (0.37)
KetoMIR		0.30	0.48 (0.37)	0.35 (0.39)	0.20 (0.45)
BHB-MIR			0.19	n.c. [1]	0.72 (0.32)
Aceton-MIR				0.16	0.61 (0.38)
Fat-protein-ratio					0.20

[1] n.c. = no convergence

Acknowledgement

This work was conducted within the COMET-Project D4Dairy (Digitalisation, Data integration, Detection and Decision support in Dairying, Project number: 872039) that is supported by BMK, BMDW and the provinces of Lower Austria and Vienna in the framework of COMET-Compentence Centers for Excellent Technologies. The COMET program is handled by the FFG, grant number 872039.

References

Belay T.K., M. Svendsen, Z.M. Kowalski and T. Adnoy, 2017. Genetic parameters of blood beta-hydroxybutyrate predicted from milk infrared spectra and clinical ketosis, and their associations with milk production traits in Norwegian Red cows. J. Dairy Sci. 100: 6298-6311.

Benedet A., A. Costa, M. De Marchi and M. Penasa, 2020. Heritability estimates of predicted blood beta-hydroxybutyrate and nonesterified fatty acids and relationships with milk traits in early-lactation Holstein cows. J. Dairy Sci. 103: 6354-6363.

Dale, L. M. and A. Werner, 2018. KetoMIR – Calculation of the ketosis risk based on MIR spectra (Model 2018). Personal communication.

Hamann H., A. Werner, L. Dale and P. Herold, 2017. Proc. of the 41st ICAR Conference, Edinburgh, United Kingdom.

Pieper, J. T. and K. Mahlkow-Nerge, 2017. In welchem Zusammenhang steht der Fett-Eiweibeta-Quotient in der Milch zur Ketose des Rindes? https://www.proteinmarkt.de/fileadmin/bilder/fachartikel/2017/FA_Rind_Nr.16_Zusammenhang_Fett-Eiweiss-Quotient_zur_Ketose.pdf

Animal breeding sustainability: the Italian Holstein experience

L. Benzoni[1], R. Finocchiaro[1], G. Niero[3], G. Invernizzi[2], G. Savoini[2], F. Galluzzo[1] and M. Cassandro[1,3]

[1]Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana
(ANAFIBJ), Via Bergamo 192, 26100 Cremona (CR), Italy
Corresponding author: lorenzobenzoni@anafi.it
[2]Dipartimento di Medicina Veterinaria e Alimentazione Animale,
Via dell'Università 6, 26900 Lodi (LO), Italy
[3]Department of Agronomy, Food, Natural resources, Animals and Environment
(DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy

Abstract

Livestock farming, especially dairy breeding, has a significant influence on environmental balance, accounting for about 50% of greenhouse gases (GHG) emissions from the primary sector. To increase the environmental sustainability of the dairy sector, even in the face of growing interests of the consumers towards this topic, a holistic approach is needed. Methane and carbon dioxide emissions have been shown to be heritable in cattle, providing the basis to apply genetic selection for their reduction. Furthermore, it is necessary to consider that GHG recording is complex, expensive and time consuming. In this context national breeding programs can provide relevant contributions. For this reason, the Italian Association of Holstein, Brown and Jersey breeders (ANAFIBJ) is working on data collection of innovative phenotypes and, in the future, to set-up routine recording in commercial dairy farms. Since 2018 ANAFIBJ, has started to record GHG data on young genotyped Italian Holstein bulls passing into the Genetic Center. For this purpose, the GreenFeed system (C-Lock Inc., Rapid City, SD) has been installed and used. In three years, a dataset of more than 11,200 phenotypic records collected on more than 200 young bulls has been set-up. Preliminary analyses showed that animals emit 223,6 g of CH4/d with a heritability (h2) of 0.396. Thanks to this experience ANAFIBJ has the intention to contribute further and set up a routine recording system for these phenotypes implementing experimental protocols to apply in commercial farms. For this purpose, Laser Methane Detector Mini (LMD, Crowcon, Abingdon, UK) is currently being tested at ANAFIBJ Genetic Center and a data collection protocol is under investigation. Once a standard protocol will be defined, individual CH4 emissions will be collected in 3,000 genotyped Italian Holstein dairy cows (some of them daughters of the young bulls recorded at the Genetic Center) distributed in 100 commercial farms throughout the country. At the Genetic Center, in addition, several phenotypes will be collected in order to better define the GHG data emission. Main biometric measures will be recorded and samples of ruminal fluid and faeces will be collected. Biological samples will be frozen and stored at -80°C, until instrumental and bioinformatic analysis. Activities in commercial farms and in experimental stations will allow to study the interaction between host and environmental microbiome, and to evaluate the reliability of faeces as a proxy of rumen sample. Furthermore, it will be possible to estimate the genetic parameters and to develop models for genetic and genomic evaluations of methane emissions.

Keywords: Greenhouse gas emissions, data-collection, Holstein, data-analysis, GreenFeed, laser methane detector.

Introduction

Agriculture and livestock sectors are recognized as important contributors to global temperature increase (Cassandro *et al*., 2013). Livestock farming, with particular regard to ruminants, is linked to GHG emissions due to enteric fermentation. Furthermore, livestock sector indirectly contributes to GHG emissions through activities related to feed production, manure spreading and storage, nitrogenous fertilisers, fossil fuels consumption and deforestation.

Methane and carbon dioxide emissions have been shown to be heritable, providing the basis to apply genetic selection for their reduction (Pickering *et al*., 2015; Lassen and Løvendahl,, 2016). Such program could be applied directly, by selecting for breath measurements, but also using indirect selection including indicator traits such as feed intake (de Haas *et al*., 2017). National breeding program can provide relevant contribution to reduce GHG emissions. The objective of this study was to present the ANAFIBJ experience and future perspective for GHG data collection and set up of a routine protocol for commercial and experimental Italian Holstein dairy farms. Furthermore, it will allow to study in deep the knowledge of the microbiome-host and environment-microbiome interactions and to evaluate the reliability of faeces as a proxy of rumen sample.

Material and methods

Animals involved in this preliminary study were young genotyped Italian Holstein bulls undergoing progeny test in the Genetic Center of the ANAFIBJ as reported by Callegaro *et al*. (2022). These young bulls had a genomic index included in the best 2% of the Italian Holstein population; they will be the future reproducers of the Italian Holstein breed. The ANAFIBJ Genetic Center is currently equipped with two instruments for individual impact recording: 1) GreenFeed *(C-Lock Inc., Rapid City, SD, USA*), and 2) The Laser Methane Detector (*LMD - Crowcon, Abingdon, UK*).

1. **The GreenFeed**, considered as an "Automated Head-Chamber System (*AHCS),* is an automated feeding station designed to measure daily CH_4 and CO_2 emissions (g/d) from ruminant's breath (Hristov *et al*., 2015). No more than 20 animals were housed in the GreenFeed box in order to ensure animal welfare and to avoid multiple animals at once. Each animal could visit the AHCS every six hours. At each visit the feed was unloaded a maximum of six times, for a total of 24 daily visits per animal. This set of traits included number of visits (NVG), carbon-dioxide daily emission (CO_2), methane daily emission (CH_4), average airflow (AIR) and average time (ATG).
2. **Laser Methane Detector Mini** (LMD). The LMD is a highly responsive, hand-held device that is pointed at an animal's nostrils and, based on infrared absorption spectroscopy, measures methane column density along the length of the laser beam (ppm*m) (Garnsworthy *et al*., 2019). This instrument is connected to a smartphone or tablet for data storing. During LMD data collection all ventilation or cooling systems inside the barn were turned off. LMD default settings were maintained with a measurement interval of 0.5 seconds (two values of CH_4 per seconds were measured). The operator was located in front of the standing

animal at a distance of 1.5 meters. Each measurement lasted 330 seconds. Each animal was recorded 3 times per day for 10 consecutive days for a total of 28 measurements per animal. The data generated make a list of CH_4 values accompanied by a unique date and time stamp and a value for the quality of the reflection of the laser beam. A single measurement consisting in a time series of CH_4 values of a single animal is called "profile". From each profile the raw mean (MEAN) of CH_4 and the mean of all peaks (P_MEAN) was calculated (Niero *et al*., 2020).

GreenFeed data were relative to 221 Italian Holstein young bulls between 171 and 541 days of age. Data have been recorded in the period between May 2018 and April 2022 and each trial lasts on average 15 days.

LMD Genetic Center phenotypic data available belong to 18 Italian Holstein young bulls between 171 and 541 days of age for a total of 483 profiles. Records with intensity less than 100 were discarded. Profiles analysis was carried as reported by Sorg *et al*., 2018. Data have been recorded in the period between January 2022 and June 2022.

Animals' biometric measures were also recorded. This set of traits included measures of body growth taken using electronic scales and stadiometers operated by qualified personnel as body weight (WEI), Body Condition Score (BCS), hearth girth (HG) and height (HT).

All young bulls were genotyped using various SNP chips resulting in 69,127 SNP. Genomic data were subsequently edited using the preGSf90 software (Aguilar *et al*., 2010), removing SNP with call rate below 0.9 and minor allele frequency below 0.05. After editing, 61,591 SNP were available.

Results

Descriptive statistics and heritability estimates for the studied traits are reported in Table 1 and Table 2. Growth traits showed the largest heritability estimates (h^2), all being above or close to 0.40. While these traits are expected to be highly heritable, the estimates appeared larger compared to those found in literature. This could be due to the relatively small sample size. Heritability estimates for the emission traits were moderate to high, ranging from 0.241 for ATG to 0.480 for CO_2.

Discussion

Heritability estimates showed substantial genetic variation for the studied traits. The CO_2 and CH_4 emissions calculated on daily basis showed high heritability with the possibility of selection and therefore possible reduction of GHG emissions. The estimated values for the heritability of CH_4 and CO_2 were higher than those found in literature (Lassen and Løvendahl, 2016; Brieder *et al*., 2019), although this could be due to the involvement of growing bulls rather than lactating cows and the limited sample size in the current study.

Table 1. Descriptive statistics (posterior means with posterior standard deviation) and heritability estimates for the traits analysed.

Trait[1]	Metric	N	Mean	SD	h^2
WEI	kg	885	309.3	77.48	0.445 (0.236)
BCS	score	849	3.0	0.33	0.512 (0.201)
HG	cm	715	157.3	14.15	0.441 (0.247)
HEI	cm	714	125.5	7.71	0.393 (0.225)
NVG	count	2817	3.9	1.71	0.360 (0.113)
CO_2	g/d	2817	6198.2	1103.88	0.480 (0.206)
CH_4	g/d	2817	223.6	51.83	0.396 (0.169)
AIR	L/s	2817	29.2	4.02	0.448 (0.088)
ATG	s	2817	329.3	87.49	0.241 (0.105)

[1] WEI: body weight; BCS: body condition score; HG: heart girth; HEI: height; NVG: number of visits at the GreenFeed; CO_2 daily carbon dioxide emissions; CH_4: daily methane emissions; AIR: average airflow at the visit; ATG: average time at the GreenFeed.

Table 2. Descriptive statistics for the traits analysed using Laser Methane Detector.

Trait[1]	Metric	N	Mean	SD
P_MEAN	ppm*m	18	53.9	30.62
MEAN	ppm*m	18	30.19	20.65

[1] P_MEAN: arithmetic mean of all peaks; MEAN: arithmetic mean of all values.

Conclusion

These preliminary results suggest that selection indexes could be implemented in order to reduce GHG emissions. ANAFIBJ is testing several experimental protocols in the Genetic Center; it is turning in a “Living Lab”. Thanks to this experience ANAFIBJ wants to contribute further and set up a routine recording system implementing experimental protocols to be applied both in experimental and commercial farms to collect a critical mass of data useful for Italian Holstein selection. Individual CH_4 emissions will be collected in 3,000 genotyped Italian Holstein dairy cows distributed homogeneously in 100 commercial farms throughout the country. Some of these cows are daughters of the young bulls recorded at the Genetic Center. In addition, other phenotypes will be collected in order to better define GHG data emissions.

Furthermore, samples of ruminal fluid, faeces and buccal swabs will be collected. Rumen fluid will be collected using the Flora Rumen Scoop (Geishauser *et al*. 2012), a ruminal probe specially designed for cattle. The first 0.5 liter of the rumen fluid will be discarded to avoid saliva contamination, and the next 0.5 liter will be retained for sampling. After each sampling the probe will be washed with water, disinfected, rinsed and dried. This procedure will allow to avoid cross-contamination of rumen fluid between animals and to avoid interference of chemical products used for disinfection. Ruminal fluid samples will be stored in individual bottles at -80°C. At the same time buccal swabs and faeces will be sampled. Buccal swabs will be collected using salivary swabs. Faeces will be obtained by rectal grab sampling using disposable gloves and will be stored in plastic bags. Ruminal fluid, buccal swabs and faeces will be submitted for metagenomic analysis. Relative abundance of OTU (Operational Taxonomic Units) will be evaluated. These activities will allow to study in deep the knowledge of the microbiome-host and environment-microbiome interaction and to evaluate the reliability of faeces as a proxy of rumen sample.

Acknowledgement

This study was supported by “Latteco2 project, sottomisura 10.2 of the PSRN Biodiversity 2020–2023” (MIPAAF. D.M. no. 465907 del 24/09/2021, project unique code 12C21004080005).

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., ***et al*****.** (2010). J. Dairy Sci. 93:2 743–752. https://doi.org/10.3168/jds.2009-2730.

Appuhamy J.A.D.R.N., France J. and Kebreab E. (2016). Glob. Chang. Biol. 22:9 3039–3056. https://doi.org/10.1111/gcb.13339.

Breider I.S., Wall E., and Garnsworthy P.C. (2019). J. Dairy Sci. 102:8 7277–7281. https://doi.org/10.3168/jds.2018-15909.

Callegaro S., Niero G., Penasa M., Finocchiaro R., Invernizzi G., Cassandro M. (2022). Italian J. Anim Sci. 21:1, 870-877, https://doi.org/10.1080/1828051X.2022.2071178.

Calus M.P.L. and Veerkamp R.F. (2011). Genet. Sel. Evol. 43:26. https://doi.org/10.1186/1297-9686-43-26

Cassandro M., Mele M., Stefanon B. (2013). Ital J. Anim Sci. 12:450–458. https://doi.org/10.4081/ijas.2013.e73.

de Haas Y., Pszczola M., Soyeurt H., Wall E., and Lassen J. (2017). J. Dairy Sci. 100:2 855–870. https://doi.org/10.3168/jds.2016-11246.

Garnsworthy, P.C., Difford, G.F., Bell, M.J., Bayat, A.R., Huhtanen, P., Kuhla, B., Lassen, J., Peiren, N., Pszczola, M., Sorg, D. (2019) Animals 9, 837. https://doi.org/10.3390/ani9100837

Geishauser,T., N. Linhart, A. Neidl, and A. Reimann (2012): J. Dairy Sci. 95, 4556-4567. https://doi.org/10.3168/jds.2012-5380.

Gerber P.J., Hristov, A.N., Henderson B., Makkar H., Oh J., ***et al*****.** (2013). Animal 7:2 220–234. https://doi.org/10.1017/S1751731113000876.

Haque M.N. (2018). J. Anim. Sci. Tech. 60:15. https://doi.org/10.1186/s40781-018-0175-7

Hristov A.N., Oh J., Giallongo F., Frederick T., Weeks, H., ***et al*****.** (2015). JoVE J. Vis. Exp. 103, 52904. https://doi.org/10.3791/52904.

Lassen J. and Løvendahl P. (2016). J. Dairy Sci. 99:3 1959–1967. https://doi.org/10.3168/jds.2015-10012.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., ***et al*****.** (2002). Proc. of the 7th WCGALP. Montpellier, France.

Niero, G.; Cendron, F.; Penasa, M.; De Marchi, M.; Cozzi, G.; Cassandro, M. (2020). Animals. 10, 606. https://doi.org/10.3390/ani10040606.

Nieuwhof G.J., van Arendonk J.A.M., Vos H., Korver S. (1992). Livestock Production Science 32:189-202. https://doi.org/10.1016/S0301-6226(12)80001-7.

Pickering N.K., Chagunda M.G.G, Banos G., Mrode R., McEwan J.C. ***et al*****.** (2015). J. Anim. Sci. 93:1 11–20. https://doi.org/10.2527/jas.2014-8302.

Sorg D., Difford G.F., Muhlbach S., Kuhla B., Swalve H. H., Lassen J., Strabel T., Pszczola M. Computers and Electronics in Agriculture. https://doi.org/10.1016/j.compag.2018.08.024.

Mid-Infrared Analyzers: Herd management milk fatty acid calibration and validation of multiple instruments

D.M. Barbano, C. Coon and M. Portnoy

Northeast Dairy Foods Research Center, Department of Food Science, Cornell University, Ithaca, NY, USA
Corresponding Author: Barbano1@aol.com

Abstract

Mid-infrared (MIR) analyzers require calibration with samples that have known reference chemistry values. Traditional milk calibration samples are individual farm, or cow milks, with reference chemistry for each sample. A more advanced procedure for making an orthogonal design (fat, protein, lactose, urea) sample calibration set (14 samples) with all-lab mean (n=8 to 10) reference chemistry was published and updated in 2020. Recently, this same sample set has been used for milk fatty acid calibration. The fatty acid reference chemistry is from gas chromatography run on the extracted fat from the ether extraction used for the fat payment test. Reference values for major individual fatty acids and those used for the most useful dairy herd management decision making (i.e., de novo, mixed origin, preformed fatty acids, and double bonds per fatty acid) are produced for this orthogonal design sample set. Reference chemistry for groups of milk fatty acids utilizes the values for only the major fatty acids (C4, 6, 8, 10, 12, 14, 16:0, C16:1, C18:0, C18:1, C18:2, C18:3) normalized to 100% and expressed as g/100 g of milk. Using only the major fatty acids will achieve better between laboratory agreement and consistency for GLC fatty acid methods. Glycerol is approximately 5.5% of the weight of milk fat. A useful quality control metric for MIR data is the sum of the de novo, mixed origin, and preformed fatty acids (g/100 g milk) divided by the fat test, and should be between 93 and 96% of the fat test in g/100 g milk. If outside this range, there is a problem with ether the MIR fat test or one or more of the MIR values for fatty acid groups. Nine MIR milk analyzers located in different regions of the US were calibrated with samples described above and then validated with a group of 8 individual farm milks from collected from different regions of the US by comparison to GLC reference chemistry on the same milks. The best agreement (g/100 g milk) for the mean of all instruments with reference chemistry was for de novo (MD -0.016 and SDD 0.028) and double bonds per fatty acid (MD 0.00 and SDD 0.01). Mixed and preformed had MD of 0.08 and -0.054 and SDD of 0.053 and 0.048, respectively. Mixed and preformed fatty acid models are more sensitive to variation in homogenizer performance than *de novo*.

Introduction

History and first field application herd management fatty acid models

The development of PLS models for mid-infrared (MIR) milk analysis of dairy herd management parameters was initiated in 2011 in a collaborative program among researchers from Cornell University (Ithaca, NY), the laboratory of St Albans Cooperative (St Albans, Vermont), and Delta Instruments, Drachten, The Netherlands. The new herd management milk analysis parameters de novo, mixed origin, and preformed fatty acid (FA) models, fatty chain length, double bonds per fatty acid (i.e., milk fat depression index), milk estimated blood NEFA were initiated and were first applied in 2012 for routine testing of producer bulk tank milks from the St Albans Cooperative (about 400 farms). Application of these same herd management models (plus models to measure milk urea, milk BHB, and milk acetone for individual cow milk testing was initiated in 2014 in a collaboration between Cornell researchers and researchers at the W. H. Miner Institute, Chazy, NY. Our first work on bulk tank milks from farms at the St Albans Cooperative found a positive relationship between higher de novo and de novo + mixed origin fatty acids with bulk tank fat and protein tests. The relationship between de novo , mixed origin, and preformed milk fatty acids with fat and protein test and fat and protein production per cow per day was the focus of joint field research study by Cornell University and W. H. Miner Institute.

The results of the field studies of the herd management milk analysis models were published in two research papers (Woolpert *et al.,* 2016 and Woolpert *et al.,* 2017). Woolpert *et al.,* 2016 and 2017 identified management practices, such as higher stall stocking density and lower feeding frequency, which were related to lower de novo FA content in bulk tank milk. Farms with lower de novo FA, on average, produced less milk fat and protein per cow per day. In addition, higher dietary EE was related to lower de novo FA content of milk. High de novo farms also had higher milk yield and fat and true protein content and yield. Management (i.e., frequent feed delivery and increased feed bunk space per cow) and dietary (i.e., adequate physically effective fiber and lower ether extract) factors that differed between high de novo and low de novo farms have been shown in earlier studies to affect ruminal function. At a constant level of milk production, the gross income per unit of milk produced was higher on the high de novo farms because of higher milk fat and true protein concentrations. Therefore, the de novo FA concentration of bulk tank milk may be an important tool for monitoring ruminal function on commercial dairy farms.

Dairy herd management milk analysis model development

Our strategy for development of herd management milk analysis PLS models was to develop milk analysis metrics that could be used more strategically and tactically to identify the causes of increases and decreases in milk fat and protein production per cow per day. With respect to herd management milk testing for diagnostic determination of why milk fat has increased or decreased, the strategy was to develop milk analysis models that provide information on milk fat production by two different milk fatty acid sources:

1. De novo synthesis of fatty acid in the mammary cells and
2. Transfer of performed fatty acid from the blood stream into mammary secretory cells.

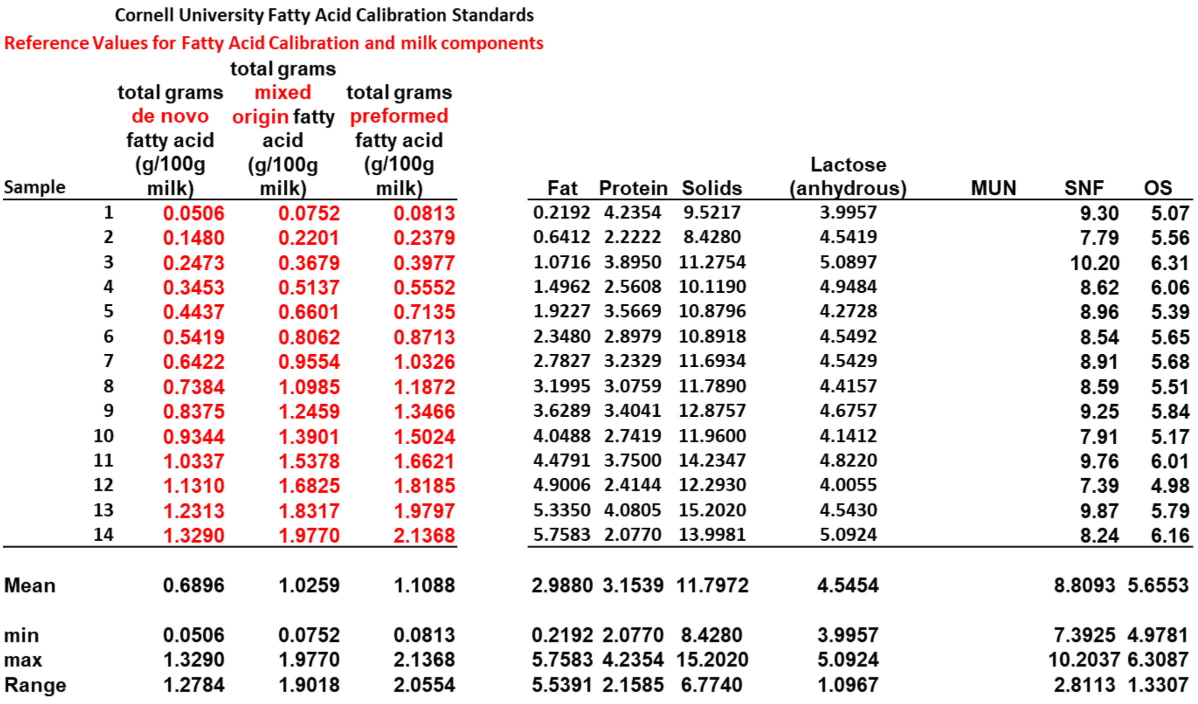
The development and modeling performance metrics (i.e., RSD values) of the PLS models for measurement of milk de novo, mixed origin, preformed, and total milk fatty acids were described by Woolpert *et al*. (2016). The development and modeling performance metrics for the fatty chain length (mean carbon number) and the milk double bonds per fatty acid models (i.e., milk fat depression index) were described by Wojciechowski and Barbano (2016). The development and modeling performance metrics for the milk estimated blood NEFA model were described by Bach *et al*. (2021).

Materials and methods

Calibration of MIR milk fatty acid models

Like all other MIR milk analysis metrics (e.g., fat, protein, solids, urea, etc.), all milk fatty prediction models need to calibrated with reference samples with known reference values. The reference values for milk fatty acids are determined by gas liquid chromatography (GLC). The sample extraction, methylation and GLC method used in our studies was described by Wojciechowski and Barbano (2016). Briefly milk fat was extracted from each sample by an ether extraction milk that is the reference for milk fat payment (AOACI, 2021, method 989.05), the formation of methyl esters was catalyzed by methanolic KOH and boron triflouide, and methyl esters of fatty acids were determined GLC as described by Wojciechowski and Barbano (2016). The calculation of milk fatty acids, fatty acid chain length and double bonds per fatty acid was described by Kaylegian *et al*. (2009). The steps of accounting for recovery of short chain fatty acids, normalization of fatty acids to 100% and removal of the impact the added methyl group on the relative proportion on each fatty acid chain length are important steps described by Kaylegian *et al*. (2009). These methods and the approach of measuring fatty acid groups (de novo, mixed origin, and performed) are applied to an orthogonal design set of MIR calibration milks (Kaylegian *et al*. 2006) that was modified to include milk urea by Portnoy *et al*. (2021). The reference values for each of the 14 milks in the calibration set for de novo, mixed origin, and preformed fatty acids are expressed as grams of fatty acid per 100 g milk, as shown in Figure 1.

Three hundred and forty sets of the calibration samples shown in Figure 1 are produced once every 4 weeks and distributed laboratories for calibration of components and milk fatty acids.



Cornell University Fatty Acid Calibration Standards

Reference Values for Fatty Acid Calibration and milk components

Sample	total grams de novo fatty acid (g/100g milk)	total grams mixed origin fatty acid (g/100g milk)	total grams preformed fatty acid (g/100g milk)	Fat	Protein	Solids	Lactose (anhydrous)	MUN	SNF	OS
1	0.0506	0.0752	0.0813	0.2192	4.2354	9.5217	3.9957		9.30	5.07
2	0.1480	0.2201	0.2379	0.6412	2.2222	8.4280	4.5419		7.79	5.56
3	0.2473	0.3679	0.3977	1.0716	3.8950	11.2754	5.0897		10.20	6.31
4	0.3453	0.5137	0.5552	1.4962	2.5608	10.1190	4.9484		8.62	6.06
5	0.4437	0.6601	0.7135	1.9227	3.5669	10.8796	4.2728		8.96	5.39
6	0.5419	0.8062	0.8713	2.3480	2.8979	10.8918	4.5492		8.54	5.65
7	0.6422	0.9554	1.0326	2.7827	3.2329	11.6934	4.5429		8.91	5.68
8	0.7384	1.0985	1.1872	3.1995	3.0759	11.7890	4.4157		8.59	5.51
9	0.8375	1.2459	1.3466	3.6289	3.4041	12.8757	4.6757		9.25	5.84
10	0.9344	1.3901	1.5024	4.0488	2.7419	11.9600	4.1412		7.91	5.17
11	1.0337	1.5378	1.6621	4.4791	3.7500	14.2347	4.8220		9.76	6.01
12	1.1310	1.6825	1.8185	4.9006	2.4144	12.2930	4.0055		7.39	4.98
13	1.2313	1.8317	1.9797	5.3350	4.0805	15.2020	4.5430		9.87	5.79
14	1.3290	1.9770	2.1368	5.7583	2.0770	13.9981	5.0924		8.24	6.16
Mean	0.6896	1.0259	1.1088	2.9880	3.1539	11.7972	4.5454		8.8093	5.6553
min	0.0506	0.0752	0.0813	0.2192	2.0770	8.4280	3.9957		7.3925	4.9781
max	1.3290	1.9770	2.1368	5.7583	4.2354	15.2020	5.0924		10.2037	6.3087
Range	1.2784	1.9018	2.0554	5.5391	2.1585	6.7740	1.0967		2.8113	1.3307

Figure 1. Milk calibration samples and reference chemistry (components and fatty acids).

Results and discussion

Between laboratory agreement: milk fatty acid validation testing

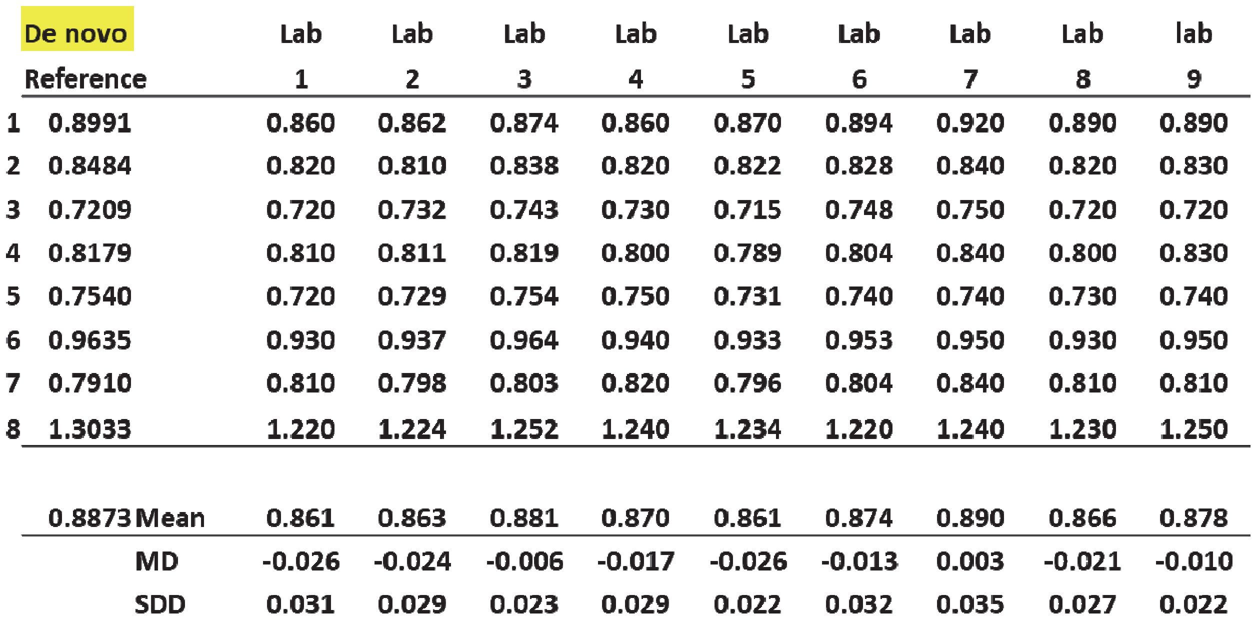
Four times per year, a set of 8 unknown individual farm milk are sent to all laboratories for performance evaluation. The farms selected for this testing are 2 farms from each of 4 regions of the US. These 8 farm milks are sent to one laboratory and they split and shipped to all laboratories for testing by MIR milk analysis and for reference testing. Tables 1 through 5 below contain an example of the results from multi-laboratory testing of the same milks. The results in the Tables below are from a group of instruments that are a mixture of Delta FTA and Delta Combi milk analyzers. Recently, additional instruments by Bentley and Foss have been using our calibration samples and in general their performance on the milk fatty acids is similar to what we have observed for the Delta instruments.

On average for this set of unknown farm milks, most MIR laboratories produced a mean estimate of the de novo fatty acids was lower than the reference chemistry. However, on an absolute basis all labs had a mean difference that was ≤ 2.93% relative of the reference chemistry mean for the sample set.

On average for this set of unknown farm milks, most MIR laboratories produced a mean estimate of mixed origin fatty acids that was higher than the reference chemistry. However, on an absolute basis all labs had a mean difference that was ≤ 8.9% relative of the reference chemistry mean for the sample set.

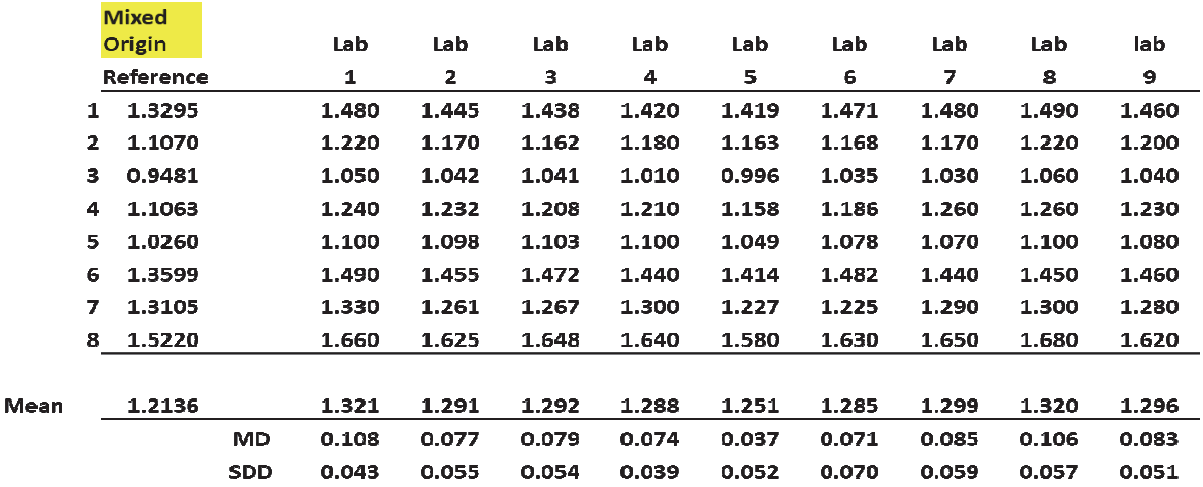
On average for this set of unknown farm milks, most MIR laboratories produced a mean estimate of performed fatty acids that was lower than the reference chemistry. However, on an absolute basis all labs had a mean difference that was ≤ 7.0% relative of the reference chemistry mean for the sample set.

Table 1. Between laboratory comparison of de novo fatty acid analysis (g/100 of milk) and calculated mean difference (MD) and standard deviation of the differences from reference chemistry for the 8 samples.



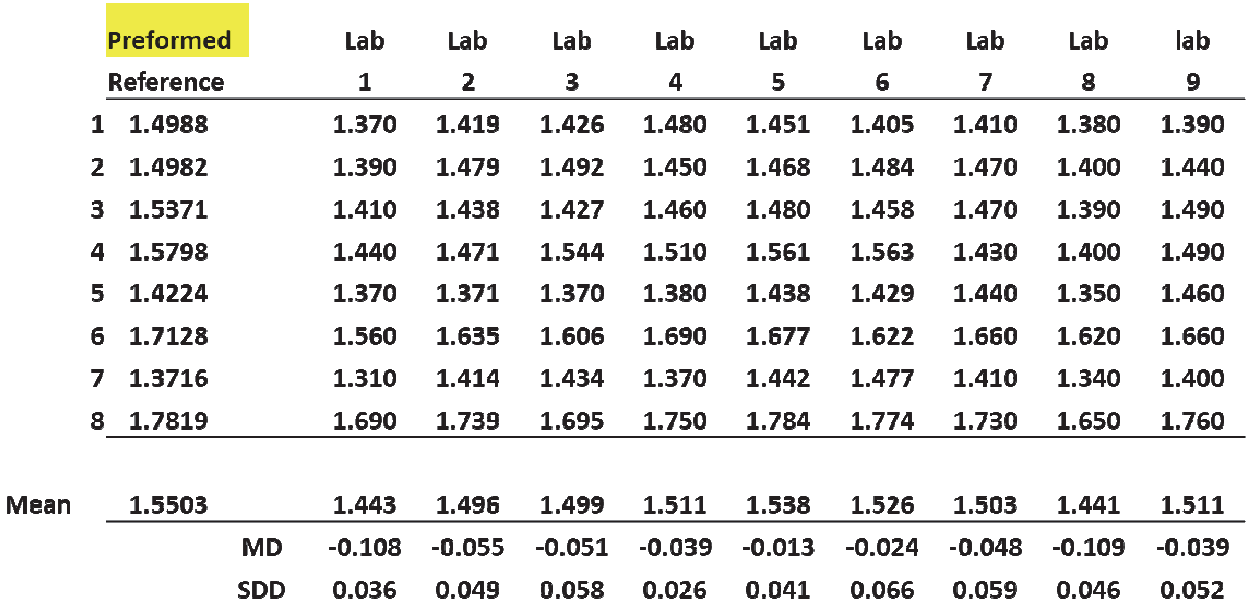
	De novo Reference		Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	lab 9
1	0.8991		0.860	0.862	0.874	0.860	0.870	0.894	0.920	0.890	0.890
2	0.8484		0.820	0.810	0.838	0.820	0.822	0.828	0.840	0.820	0.830
3	0.7209		0.720	0.732	0.743	0.730	0.715	0.748	0.750	0.720	0.720
4	0.8179		0.810	0.811	0.819	0.800	0.789	0.804	0.840	0.800	0.830
5	0.7540		0.720	0.729	0.754	0.750	0.731	0.740	0.740	0.730	0.740
6	0.9635		0.930	0.937	0.964	0.940	0.933	0.953	0.950	0.930	0.950
7	0.7910		0.810	0.798	0.803	0.820	0.796	0.804	0.840	0.810	0.810
8	1.3033		1.220	1.224	1.252	1.240	1.234	1.220	1.240	1.230	1.250
	0.8873	Mean	0.861	0.863	0.881	0.870	0.861	0.874	0.890	0.866	0.878
		MD	-0.026	-0.024	-0.006	-0.017	-0.026	-0.013	0.003	-0.021	-0.010
		SDD	0.031	0.029	0.023	0.029	0.022	0.032	0.035	0.027	0.022

Table 2. Between laboratory comparison of mixed origin fatty acid analysis (g/100 of milk) and calculated mean difference (MD) and standard deviation of the differences from reference chemistry for the 8 samples.



	Mixed Origin Reference		Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	lab 9
1	1.3295		1.480	1.445	1.438	1.420	1.419	1.471	1.480	1.490	1.460
2	1.1070		1.220	1.170	1.162	1.180	1.163	1.168	1.170	1.220	1.200
3	0.9481		1.050	1.042	1.041	1.010	0.996	1.035	1.030	1.060	1.040
4	1.1063		1.240	1.232	1.208	1.210	1.158	1.186	1.260	1.260	1.230
5	1.0260		1.100	1.098	1.103	1.100	1.049	1.078	1.070	1.100	1.080
6	1.3599		1.490	1.455	1.472	1.440	1.414	1.482	1.440	1.450	1.460
7	1.3105		1.330	1.261	1.267	1.300	1.227	1.225	1.290	1.300	1.280
8	1.5220		1.660	1.625	1.648	1.640	1.580	1.630	1.650	1.680	1.620
Mean	1.2136		1.321	1.291	1.292	1.288	1.251	1.285	1.299	1.320	1.296
		MD	0.108	0.077	0.079	0.074	0.037	0.071	0.085	0.106	0.083
		SDD	0.043	0.055	0.054	0.039	0.052	0.070	0.059	0.057	0.051

Table 3. Between laboratory comparison of preformed fatty acid analysis (g/100 of milk) and calculated mean difference (MD) and standard deviation of the differences from reference chemistry for the 8 samples.



	Preformed Reference		Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	lab 9
1	1.4988		1.370	1.419	1.426	1.480	1.451	1.405	1.410	1.380	1.390
2	1.4982		1.390	1.479	1.492	1.450	1.468	1.484	1.470	1.400	1.440
3	1.5371		1.410	1.438	1.427	1.460	1.480	1.458	1.470	1.390	1.490
4	1.5798		1.440	1.471	1.544	1.510	1.561	1.563	1.430	1.400	1.490
5	1.4224		1.370	1.371	1.370	1.380	1.438	1.429	1.440	1.350	1.460
6	1.7128		1.560	1.635	1.606	1.690	1.677	1.622	1.660	1.620	1.660
7	1.3716		1.310	1.414	1.434	1.370	1.442	1.477	1.410	1.340	1.400
8	1.7819		1.690	1.739	1.695	1.750	1.784	1.774	1.730	1.650	1.760
Mean	1.5503		1.443	1.496	1.499	1.511	1.538	1.526	1.503	1.441	1.511
		MD	-0.108	-0.055	-0.051	-0.039	-0.013	-0.024	-0.048	-0.109	-0.039
		SDD	0.036	0.049	0.058	0.026	0.041	0.066	0.059	0.046	0.052

We have observed that the mixed and preformed milk fatty acids tend to deviate in opposite direction relative reference chemistry values. The mixed origin and preformed fatty acid models are more sensitive to variation in homogenizer performance than de novo fatty predictions or prediction of the main milk components.

On average for this set of unknown farm milks, most MIR laboratories produced a mean estimate of fatty acid chain length that was lower than the reference chemistry. However, on an absolute basis all labs had a mean difference that was ≤ 0.92% relative of the reference chemistry mean for the sample set.

On average for this set of unknown farm milks, most MIR laboratories produced a mean estimate of mean fatty acid unsaturation that was very close to the reference chemistry. On an absolute basis all labs had a mean difference that was ≤ 4.4% relative of the reference chemistry mean for the sample set. The between lab agreement on this parameter

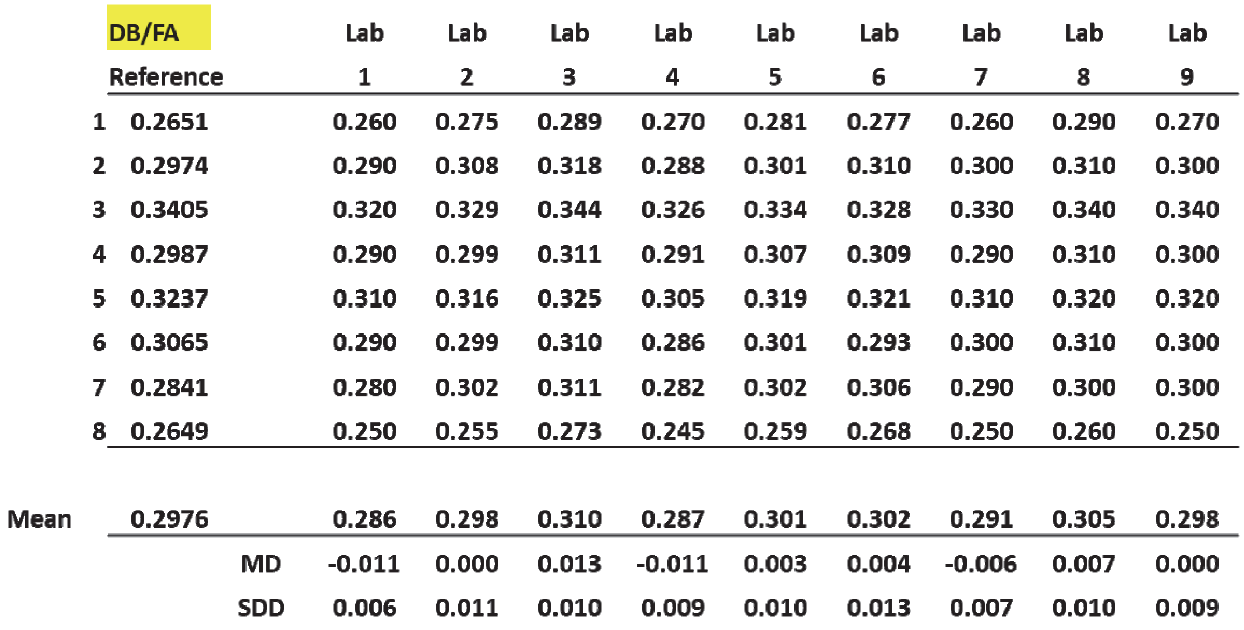
Recognizing when there is laboratory problem with MIR fatty acid data

- **Bulk tank producer milks**: The sum of de novo, mixed, and preformed fatty acids (g/100 g milk) should be about 94.5% of the fat test (g/100 g milk). This will vary from farm-to-farm, but all values should be between 93 and 96%. If sum of de novo, mixed and preformed as a % of fat test increases or decreases with change in fat concentration for a population of farms, the error is in the slope setting for the prediction of total fat, not the fatty acid testing.
- If all tests are lower than 93 or higher than 96% then it could be bias error in fat test, or the wrong reference chemistry has been used to adjust the slope and intercept on one of the 3 fatty acid measures.

Table 4.. Between laboratory comparison of fatty acid chain length analysis (carbons per fatty acid) and calculated mean difference (MD) and standard deviation of the differences from reference chemistry for the 8 samples.

	CL Reference	**Lab 1**	**Lab 2**	**Lab 3**	**Lab 4**	**Lab 5**	**Lab 6**	**Lab 7**	**Lab 8**	**Lab 9**
1	14.7434	14.63	14.76	14.80	14.72	14.65	14.65	14.67	14.76	14.76
2	14.7429	14.64	14.78	14.79	14.69	14.61	14.69	14.71	14.77	14.78
3	14.8803	14.75	14.85	14.91	14.83	14.76	14.73	14.82	14.88	14.88
4	14.7634	14.64	14.72	14.76	14.68	14.64	14.65	14.64	14.77	14.73
5	14.7897	14.67	14.75	14.78	14.71	14.66	14.67	14.73	14.76	14.78
6	14.8062	14.61	14.74	14.77	14.69	14.63	14.61	14.70	14.77	14.77
7	14.7861	14.67	14.79	14.83	14.73	14.68	14.69	14.73	14.76	14.82
8	14.4498	14.32	14.38	14.46	14.37	14.25	14.32	14.32	14.43	14.47
Mean	14.7452	14.616	14.721	14.763	14.678	14.610	14.626	14.665	14.738	14.749
	MD	-0.129	-0.024	0.017	-0.068	-0.135	-0.119	-0.080	-0.008	0.004
	SDD	0.029	0.039	0.032	0.028	0.037	0.043	0.035	0.023	0.028

Table 5. Between laboratory comparison of mean fatty acid unsaturation (double bonds per fatty acid) and calculated mean difference (MD) and standard deviation of the differences from reference chemistry for the 8 samples.



	DB/FA Reference		Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	Lab 9
1	0.2651		0.260	0.275	0.289	0.270	0.281	0.277	0.260	0.290	0.270
2	0.2974		0.290	0.308	0.318	0.288	0.301	0.310	0.300	0.310	0.300
3	0.3405		0.320	0.329	0.344	0.326	0.334	0.328	0.330	0.340	0.340
4	0.2987		0.290	0.299	0.311	0.291	0.307	0.309	0.290	0.310	0.300
5	0.3237		0.310	0.316	0.325	0.305	0.319	0.321	0.310	0.320	0.320
6	0.3065		0.290	0.299	0.310	0.286	0.301	0.293	0.300	0.310	0.300
7	0.2841		0.280	0.302	0.311	0.282	0.302	0.306	0.290	0.300	0.300
8	0.2649		0.250	0.255	0.273	0.245	0.259	0.268	0.250	0.260	0.250
Mean	0.2976		0.286	0.298	0.310	0.287	0.301	0.302	0.291	0.305	0.298
		MD	-0.011	0.000	0.013	-0.011	0.003	0.004	-0.006	0.007	0.000
		SDD	0.006	0.011	0.010	0.009	0.010	0.013	0.007	0.010	0.009

- **Individual cow milks**: The sum of de novo, mixed, and preformed fatty acids should be about 94.5% of the fat test. This will vary from cow-to-cow. The errors discussed on the previous slide for bulk tank milk can have the same impact on individual cow milk tests.
- The PLS fatty acid herd management models were developed for bulk tank milks and for milks from individual cows in positive energy balance. Thus, these fatty acid models will not work well on milks from cows at less than 5 days in lactation. Very early lactation milks will give values for the sum of the fatty acid (g/100 g milk) that exceeds the fat test in g/100 g milk. These samples are outside the scope of the ability of the current models.
- **Herd management application of milk fatty data.** This is a discussion that is beyond the scope of this presentation, but we have presented data from field studies and provided examples of how to interpret the fatty acid data from bulk tank milks. Those examples are provided in a series of papers presented at the annual Cornell Dairy Conference (Barbano *et al* 2014,2017, 2018, 2019 ; Barbano and Mellili 2016).

Conclusion

Milk fatty acid analysis is a useful tool for dairy herd management. Most dairy herd management milk fatty acid analysis in the USA has been applied to bulk tank and tanker load samples from individual farms in milk payment testing laboratories. Today bulk tank milk is tested on virtually every pick up and for large farms every tanker load of milk is tested daily. Results from this testing are usually posted within 36 h of sample collection. This has provided a valuable resource for nutrition management on dairy farms in the USA and dairy nutritionists have rapidly improved their skills for interpretation of the data. Monitoring milk fatty acid composition across time on the same farm (g/100 g milk) has proved very useful in management of dairy rations.

Individual cow milk testing is still a challenge. In general, the frequency of milk sampling and testing of milks from individual cows is too low to provide useful information for farm management. The interpretation of the data is more complex because there are normal and systematic changes in milk fatty acid composition with stage of lactation and these factors need to be considered when interpreting milk fatty acid data from individual cows. In the long- term future, the dairy industry needs to strive to achieve milk testing hardware innovations that allow individual cow milk analysis to occur in real-time on the farm during milking. If that is achieved, monitoring and management of individual cow health will advance rapidly.

References

Bach, K. D., D. M. Barbano, and J. A. A. McArt. 2021. The relationship of excessive energy deficit with milk somatic cell score and clinical mastitis. J. Dairy Sci. J. Dairy Sci. 104:715–727.

Barbano, D. M., C. Melilli, and T. Overton. 2014. Advanced use of FTIR spectra of milk for feeding and health management. Pages 105 – 113 in Proceedings of the Cornell Dairy Nutrition Conference. Department of Animal Science, Cornell University, Ithaca, NY.

Barbano, D. M. and C. Melilli. 2016. New Milk Analysis Technologies to Improve Dairy Cattle Performance. Pages 61 to 73 in Proceedings of the Cornell Dairy Nutrition Conference, Department of Animal Science, Cornell University, Ithaca, NY.

Barbano, D. M., C. Melilli, H. Dann, and R. J. Grant. 2017. Infrared Milk Fatty Acid Analysis: Experience in the Field for Farm Management. Page 159 to 174 in the 2017 Proceedings of the Cornell Nutrition Conference for Feed Manufacturers. October 17 to 19, 2017. Syracuse, NY.

Barbano, D. M. H. M. Dann, C. Melilli, and R. J. Grant. 2018. Milk analysis for dairy herd management: today and in the future. Page 102 to 115 in the 2018 Proceedings of the Cornell Nutrition Conference for Feed Manufacturers. October 16 to 18, 2018. Syracuse, NY.

Barbano, D.M., H. Dann, A. Pape, C. Melilli, and R. Grant. 2019. Herd Management Milk Analysis: Jersey versus Holstein and Between Lab Agreement of Results. Page 101 to 117 in the 2019 Proceedings of the Cornell Nutrition Conference for Feed Manufacturers. October 22 to 24, 2019. Syracuse, NY.

Dann, H. M., D. M. Barbano, A. Pape, and R. J. Grant. 2018. Mid-infrared milk testing for evaluation of health status in dairy cows. Page 106 to 128 in the 2018 Proceedings of the Cornell Nutrition Conference for Feed Manufacturers. October 16 to 18, 2018. Syracuse, NY.

Kaylegian, K. E., G. E. Houghton, J. M. Lynch, J. R. Fleming, and D. M. Barbano. 2006. Calibration of Infrared Milk Analyzers: Modified Milk Versus Producer Milk. J. Dairy Sci. 89:2817–2832.

Kaylegian, K. E., D. A. Dwyer, J. M. Lynch, D. E. Bauman, J. R. Fleming, and D. M. Barbano. 2009. Impact of fatty acid composition on the accuracy of mid-infrared fat analysis of farm milks. J. Dairy Sci. 92:2502–2513.

Portnoy, M. C. Coon, and D. M. Barbano. 2021. Infrared milk analyzers: Milk urea nitrogen calibration. J. Dairy Sci. 04:7426–7437.

Wojciechowski, K. L, and D. M. Barbano. 2016. Prediction of fatty acid chain length and unsaturation of milk fat by mid-infrared milk analysis. J. Dairy Sci. 99:8561-8570.

Woolpert, M. E., H. M. Dann, K. W. Cotanch, C. Melilli, L. E. Chase, R. J. Grant, and D. M. Barbano. 2016. Management, nutrition, and lactation performance are related to bulk tank milk de novo fatty acid concentration on northeastern US dairy farms. J. Dairy Sci. 99:8486-8497.

Woolpert, M.E., H. M. Dann, K. W. Cotanch, C. Melilli, L. E. Chase, R. J. Grant, and D. M. Barbano. 2017. Management practices, physically effective fiber, and ether extract are related to bulk tank milk de novo fatty acid concentration on Holstein dairy farms. J. Dairy Sci. 100:5097–5106.

The role of recording and evaluating calf traits for improved sustainability

M.M. Axford[1,2,3*], M. Khansefid[1], A.J. and J.E. Pryce[1,2]

[1]Agriculture Victoria, AgriBio, Bundoora, Victoria 3083, Australia
[2]School of Applied Systems Biology, La Trobe University,
Bundoora, Victoria 3083, Australia
[3]DataGene Ltd, 5 Ring Road, Bundoora, Victoria 3083, Australia
Corresponding Author: maxford@datagene.com.au

Abstract

In a grazing system, about one-quarter of the costs associated with rearing dairy heifers are incurred between birth and weaning. In addition, the risk of mortality is considerably higher during this period which influences farm profitability as well as consumer attitudes towards the dairying industry. Genetic tools, such as estimated breeding values, can contribute towards improved outcomes for calves and easier calving routines for farmers. Stillbirth and heifer livability are two examples of calf traits. Acknowledging the astute ability of calf rearers to differentiate between calves that are easy or difficult to rear, a farmer-scored trait of calf vitality is under investigation. This study describes variation in calf vitality scores amongst a group of Australian Holstein herds. .

Keywords: calf vitality, calf health, dairy cattle breeding.

Background

In the Australian dairy context, sustainability is described as enhancing farmer livelihoods, improving the wellbeing of people, providing best care for all our animals and reducing environmental impact (Dairy Australia, 2020). Heifers often comprise around one-third of the total herd and are costly to rear. In a recent analysis, the cost of heifer rearing was estimated to be between AU$1190 (75% grazing) and AU$1718 (zero grazing) per heifer, of which $315 is incurred between birth to weaning in both systems (Shannon *et al*. 2022). Reducing stillbirth, calf morbidity and mortality can improve profitability by spreading costs over a larger number of healthy animals to be sold or retained as herd replacements. At the same time, farmers have an obligation to provide the best care for calves.

There is strong evidence that breeding values are useful genetic tools to improve the production and longevity of dairy cows (Veerkamp and van Pelt 2020) and thereby contribute to improved sustainability. In a specific calf example, the genetic trend for calving ease in Holstein bulls demonstrates a 1% increase in normal or easier calvings for every 5 years since 2000 in Australia. Easier calving is beneficial to both cows and their calves (Eaglen *et al*. 2011; Eaglen *et al*. 2013; Murray and Leslie 2013). A further example is found in the heritability of the calf survival trait that is estimated to be 5% for survival to 10 months in the UK (Winters 2019), 0.72% for survival to 18 months in the USA (Neupane *et al*. 2021), 0.60% for survival to 365 days in the USA (Gonzalez-Peña *et al*. 2019) and 4.2% for survival to first calving in Denmark (Fuerst-Waltl and Sørensen 2010) in Holsteins. Finally, stillbirth heritability estimates are also varied but are generally higher (1-12%) as summarised by Cole *et al*. (2007). It seems clear that breeding for calf traits can also contribute to sustainability.

In the analysis of stillbirth, calf survival and calf health traits, there is often insufficient data for a robust analysis, especially in countries where data recording is not enforced through legislation or strict quality schemes. So, the question becomes what calf trait recording initiatives could be implemented to increase the quality and quantity of calf health data that improves the genetic evaluation of calf traits and ultimately lifts calf performance?

Introducing Calf Vitality Score

A research project is underway to better understand the genetic variation in stillbirth, calf morbidity and mortality in Australian dairy calves. The project has already recruited more than fifty herds, of which six herds formed an early pilot group to test data collection processes. Several farmers in the early pilot group requested the opportunity to record something more than stillbirth, morbidity and mortality. They wanted to record calves that were exceptionally easy to rear in contrast to calves that were dull, sluggish and difficult to teach to drink. Using Australian colloquial language, calf rearers describe them as calves that were 'Rippers' or 'Duds'. As a result, the data collection was broadened to include a subjective calf vitality score.

Subjective trait measures are successfully used in the genetic evaluation systems of various countries. An example is the farmer scoring of milking speed, temperament and likeability in Australian dairy cattle with heritability estimates of 0.18-0.26 in Holstein and 0.25-0.29 in Jerseys (Beard and Jones 1991; Beard 1993; Visscher and Goddard 1995). Beard and Jones (1991) reported that Likeability, as a 'catch all' trait is a useful trait in models that predict early survival in animals without complete phenotypes. Despite the subjectivity of their scoring, these workability traits are influential in current breeding indexes (Byrne *et al*. 2016).

Given there is already familiarity with workability traits, we used a similar scoring scale as the base for a new Calf Vitality Score and modified to include aspects of behavioural scoring in beef cattle (Parham *et al*. 2019; Ceballos *et al*. 2021) and health scoring in humans (Streiner 2015). Key features of a successful scoring system will likely have the following characteristics:

- Clear terminology that is easily understood by observers.
- Simple definitions that can be memorized.
- Categories that are sufficiently different from each other.
- 3-5 categories.

The four-category calf vitality score is described in figure 1 where A-Ripper/Great, B-Good, C-Dull/Sick/Dud, D-Dead. The visual representation of the score is presented to support the needs of people with a range of learning styles. The consensus of the pilot group was that the calves should be scored for calf vitality within the first month of life.

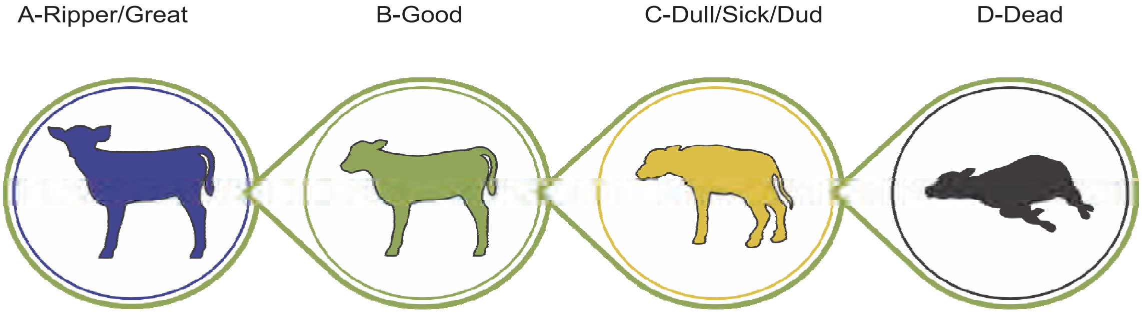


Figure 1. Calf Vitality Score.

Materials and methods

Six commercial dairy herds located in South-Eastern Australia were recruited in Spring 2020 as pilot participants, of which three recorded calf vitality scores. Farmers were asked to record calf identity, pedigree, calving dates, calf fate (stillbirth), calving ease, calf size, health and treatment events to weaning, deaths, deformed calves and calf vitality score. All stillborn calves were recorded, regardless of breed or gender. Health and vitality records were collected for dairy heifers and bulls kept for rearing. It is important to note that the calf vitality score was co-developed, with farmers, during the collection period.

Participation in the project was voluntary. Participants were invited to participate if they were passionate about calf health, willing to:

1. collect the required data;
2. provide feedback and information to the project, had calves that are predominantly sired by an AI sire and used Holstein, Jersey or Red Breed sires.

Data was supplied in a wide variety of paper-based and electronic formats but compiled into a Microsoft Access database (Microsoft) and then analysed in R Studio (RStudio Team 2021). Health and treatment events were included in this analysis so that a health event could be a diagnosis (such as scours) and/or treatment for an illness (such as treatment for scours), as recorded by the farmer. Herds were predominantly Holstein and were located within 15 km of each other meaning that calves were likely to be exposed to similar weather conditions, although management practices are expected to vary. Subsequently, we calculated summary statistics such as means and standard deviations for each herd and repeated for the overall group..

Results

There were 228, 127 and 171 recorded calves in each of 3 herds with calf vitality scores totalling 506 calves born between July and October 2020. As shown in Table 1, the mean mortality for each herd was 11, 35 and 2% including stillbirths. Stillbirths were 8, 31 and 2% respectively.

The most recorded health events were scours (37% of calves) and pneumonia (1% of calves). Calf vitality scores were recorded for 452 calves with means and ranges of Score A 32% (15-77), Score B 56% (0-73), Score C 8% (0-24) and Score D 3% (0-4). No animals with a Score A had a recorded health event. The prevalence of health events increased with subsequent scores as shown in Figure 2. Fifty-seven percent of calves with a Score B had at least one health event and this increased to 67% and 87% for Scores C and D.

Table 1. Mortality, stillbirth and calf vitality scores in 3 pilot herds.

	Herd A	Herd B	Herd C	Overall
Mortality				
Observations	228	127	151	506
Mean (±SD)	0.114	0.354	0.020	0.146 ± 0.354
Stillbirth				
Observations	228	127	151	506
Mean (±SD)	0.079	0.307	0.020	0.119 ± 0.324
Vitality				
Observations	203	98	151	452
Score A	0.241	0.765	0.146	0.323
Score B	0.714	0	0.728	0.564
Score C	0	0.235	0.086	0.080
Score D	0.044	0	0.040	0.033

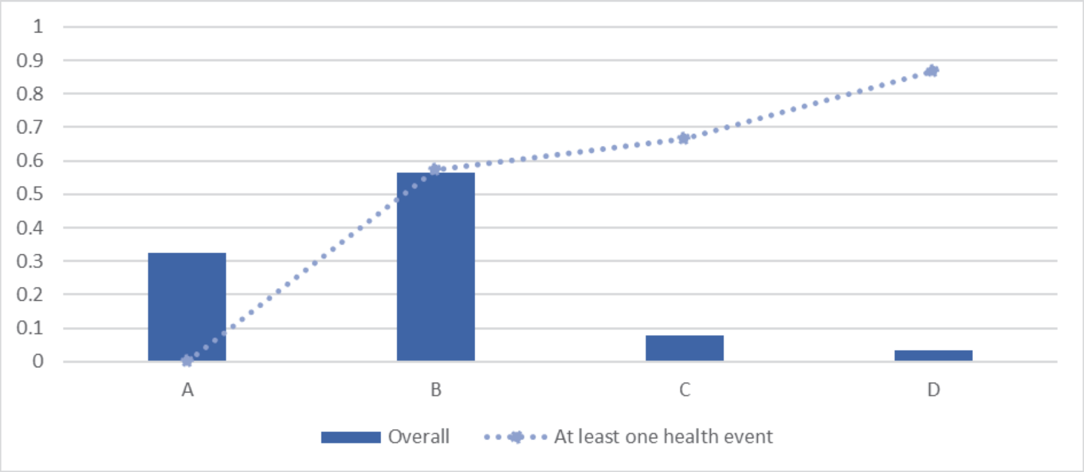


Figure 2. Mean proportions of calf vitality scores and recording of illness in 3 pilot herds.

Discussion

Farmers in the pilot group were co-developing the calf vitality scoring system during the first season of collection. The scale, descriptions, timing of scoring and recording spreadsheets were developed as the scoring was taking place. This is the likely reason for different scoring patterns between herds, as indicated by the range in mean scores. However, early indications suggest that the scoring system made sense as there were zero calves with Score A that also had a recorded health event. They were accurately recorded as 'Rippers'. In contrast, 87% of calves with Score D had a recorded health event.

While this is an early indicator that the scoring system could be useful, it also suggests that the variation in vitality scores could be so strongly linked to health events that recording health events alone could be of close to equal value. Herd recording systems are already in place for recording health events in dairy heifers. Relying on health event recording would avoid further software development tasks that would be required to systematically collect calf vitality scores nationwide. However, current recording systems typically exclude records for beef cross heifers and bulls. From a sustainability perspective, adequately managing and monitoring the health and welfare of these calves are equally as important as dairy replacements.

Following this pilot, the calf vitality score descriptions, images, recording sheets and timing of scoring were refined for use in subsequent calvings and in a much larger group of herds. For example, farmers reported that the best time to score was after the first week of life but it needed to be done before 4 weeks when differences became less obvious. Secondly, images were drawn and refined to provide a visual representation of each score. Further analysis of the larger collections that followed in 2021 and 2022 are expected to provide further insight to the value of calf vitality scoring.

Conclusion

Results from the pilot herds show that it is possible to score calf vitality as an additional piece of information in calf recording routines. Early indications suggest that calf vitality scoring is helpful in increasing engagement with farmers that can lead to improved calf health data recording. Higher quality and quantity of calf health records, for complete cohorts, will be useful to improve the benchmarking of sustainability metrics like mortality and morbidity as well as producing new genetic breeding tools. However, it is too soon to assess the value of calf vitality as an independent score. Continued collection of calf vitality scores in 2021 and 2022 has provided a larger volume of data which will be later analysed to determine the value of vitality scoring in genetic predictions of calf health traits, over and above conventional health recording.

Acknowledgement

This research has been funded by DairyBio, Agriculture Victoria and DataGene, with support from Zoetis. The authors acknowledge the significant effort of participating farmers as well as herd recording software organisations, herd recording centres and genomic service providers for their contributions.

References

Beard K (1993) Genetic evaluation for milking speed, temperament, likability and survival in Australia. In 'Interbull meeting. Arhus, Denmark'. Available at file:///C:/Users/vick3rw/Downloads/165-Article%20Text-329-1-10-20120221.pdf [Accessed 25/02/2021].

Beard K, Jones, L (1991) 'Traits affecting survival of cows in australian dairy herds, 9th Conference Australian Association of Animal Breeding and Genetics ' Melbourne, Australia.

Byrne TJ, Santos, BFS, Amer, PR, Martin-Collado, D, Pryce, JE, Axford, M (2016) New breeding objectives and selection indices for the Australian dairy industry. J Dairy Sci 99, 8146-8167.

Ceballos MC, Góis, KCR, Sant'Anna, AC, Wemelsfelder, F, Paranhos da Costa, M (2021) Reliability of qualitative behavior assessment (QBA) versus methods with predefined behavioral categories to evaluate maternal protective behavior in dairy cows. Applied animal behaviour science 236, 105263.

Cole JB, Wiggans, GR, VanRaden, PM (2007) Genetic Evaluation of Stillbirth in United States Holsteins Using a Sire-Maternal Grandsire Threshold Model. J Dairy Sci 90, 2480-2488.

Dairy A (2020) Dairy Sustainability Report 2020. Dairy Australia, Melbourne, Australia. Availableat https://www.dairyaustralia.com.au/about/collaboration-partners/australian-dairy-sustainability-framework#.YoLu3vhByUk.

Eaglen SAE, Coffey, MP, Woolliams, JA, Mrode, R, Wall, E (2011) Phenotypic effects of calving ease on the subsequent fertility and milk production of dam and calf in UK Holstein-Friesian heifers. J Dairy Sci 94, 5413-5423.

Eaglen SAE, Coffey, MP, Woolliams, JA, Wall, E (2013) Direct and maternal genetic relationships between calving ease, gestation length, milk production, fertility, type, and lifespan of Holstein-Friesian primiparous cows. J Dairy Sci 96, 4015-4025.

Fuerst-Waltl B, Sørensen, MK (2010) Genetic analysis of calf and heifer losses in Danish Holstein. J Dairy Sci 93, 5436-5442.

Gonzalez-Peña D, Vukasinovic, N, Brooker, JJ, Przybyla, CA, DeNise, SK (2019) Genomic evaluation for calf wellness traits in Holstein cattle. J Dairy Sci 102, 2319-2329.

Microsoft 'Microsoft Access for Microsoft 365.'

Murray CF, Leslie, KE (2013) Newborn calf vitality: Risk factors, characteristics, assessment, resulting outcomes and strategies for improvement. Vet J 198, 322-328.

Neupane M, Hutchison, JL, Van Tassell, CP, VanRaden, PM (2021) Genomic evaluation of dairy heifer livability. J Dairy Sci 104, 8959-8965.

Parham JT, Tanner, AE, Wahlberg, ML, Grandin, T, Lewis, RM (2019) Subjective methods to quantify temperament in beef cattle are insensitive to the number and biases of observers. Applied animal behaviour science 212, 30-35.

RStudio Team (2021) 'RStudio: Integrated Development Environment for R.' (PBC, Boston, MA:

Shannon P, Musson, J, Scott, BA, 2022. How much does it really cost to rear a heifer? Dairy Australia,

Streiner DL (Eds GR Norman, J Cairney (2015) 'Health measurement scales: a practical guide to their development and use.' (New York, New York : Oxford University Press:

Veerkamp R, van Pelt, ML (2020) Advances in dairy cattle breeding to improve longevity. In 'Advances in breeding of dairy cattle.' (Eds Jvd Werf, J Pryce.) pp. 337-353. (Cambridge, England : Burleigh Dodds Science Publishing:

Visscher PM, Goddard, ME (1995) Genetic Parameters for Milk Yield, Survival, Workability, and Type Traits for Australian Dairy Cattle. J Dairy Sci 78, 205-220.

Winters M, 2019. Calf Survival Index. Agriculture and Horticulture Development Board, Warwickshire, UK. 2.

Using dairy herd improvement data to monitor herd sustainability

D. Warner, J. Philion, R. Lacroix, D. E. Santschi and S. Adam

Lactanet, 555, boul. des Anciens-Combattants, Sainte-Anne-de-Bellevue,
H9X 3R4, Québec, Canada
Corresponding Author: dsantschi@lactanet.ca

Abstract

Milk recording data have the potential to provide a dynamic picture of the overall herd status and management practices. We developed a dairy management tool to assess and monitor herd sustainability based on 10 routinely collected DHI indicators linked to key aspects of dairy management over the life span of a dairy cow. The indicators were selected by a committee of dairy industry experts and are related to longevity and culling (% cows in lactation ≥3; involuntary culling; cow mortality), feeding and production (% cows with milk urea nitrogen <5 or >18 mg/dL; Transition Cow Index; a management score index to evaluate how well the genetic potential of cows to produce milk is expressed), heifer management (calf mortality; age at first calving), and health (% cows with BHB >0.20 mM at first test; % cows with subclinical mastitis as SCC >200,000) at herd level (12-month rolling averages except for early lactation). Indicators were aggregated to a composite herd sustainability index (SI) and aimed to benchmark the overall herd sustainability of a herd relative to its peers and highlight specific areas with opportunities for improvements. The herd SI and benchmarks are computed three times per year and made available to dairy producers and advisor services via customized reports. Preliminary analyses were conducted with 2,608 dairy herds across Quebec and New Brunswick, Canada, to validate the herd SI with herd performance and profitability (12-month test date averages for 2020), and farm management practices collected through an online survey in 2020. Overall, with increasing herd SI, the herd performance improved for production (+8.9 kg/cow and day; top 10% versus bottom 10% herds for all reported results), herd longevity (−36% involuntary culling and −74% cow mortality), heifer management (−62% calf mortality), reproduction (−36 days open and −37 days for calving interval), for health parameters (−51% potential subclinical mastitis; −71% potential hyperketonemia), and herd profitability (+26% milk value and +$93,810 in operation margin). A descriptive study of association with farm management practices on 2,143 dairy farms highlighted some important aspects of calf, heifer, dry cow, and milking management. In particular, improving the stall surface and housing comfort, colostrum management, ventilation, and udder and hoof health were key aspects to improve the herd SI. In conclusion, milk recording data can be useful to assess and monitor herd sustainability and can help dairy producers adopt best management practices to improve the production performance and herd profitability.

Keywords: Decision support, DHI, Benchmarking, Sustainability

Introduction

Sustainability is a global concept addressing the balance and long-term viability of a system. At the herd level, a sustainable herd would be a herd in an optimal balance between productivity, profitability, and longevity of the cows in the herd. This would therefore add-up to an overall performing herd with minimal incidence of disease and allowing for a good work-life balance of the producers. Using a similar approach by integrating different aspects of sustainability, Ryan *et al*. (2016) showed that the most profitable herds were also the ones performing best from an environmental sustainability perspective.

Due to the multidimensional nature of sustainability, a composite index for sustainability was created. A proof-of-concept with a comprehensive validation of the robustness of the composite index was published elsewhere (Warner *et al*., 2020). The results showed that herds could be evaluated relative to their peers while maintaining the ability to intervene on specific areas with opportunities for improvement through the individual indicators.

Our objective was to evaluate the possibility to use routinely collected DHI information to calculate herd sustainability, and then compare this ranking to performance and profitability rankings.

Material and methods

To select the proper criteria to be included in the index, and with the objective of encouraging future adoption of the SI, we conducted a consultation with a committee of industry and research experts from various fields including nutrition and management, welfare, economics, milk recording. This committee had the precise mandate to come up with a list of routinely collected dairy herd improvement (DHI) variables that could be used to identify key parameters that contribute to overall herd sustainability. In addition, the committee had to ensure that these indicators would cover the full life cycle at the farm (birth to culling) without overlapping among chosen criteria.

Sustainability Indicators

A total of 10 indicators was chosen and grouped in four categories as detailed below.

Longevity and culling

1. *Cows in 3rd lactation and over (%):* percentage of cows in the herd that started at least a 3rd lactation in the last 12 months before the index is calculated.
2. *Involuntary Culling (%):* percentage of cows removed from the herd in the last 12 months for reasons classified as involuntary culling (reproduction, feet and legs, udder health, injury, …)
3. *Cow Mortality Rate (%):* percentage of cows removed from the herd for mortality during the last 12 months prior to the calculation of the index.

Feeding and production

1. *Milk Urea Nitrogen (MUN) <5 or >18 mg N/dL (%):* average percentage of cows tested individually for MUN that had a result of less than 5 mg of nitrogen per dL OR more than 18 mg per dL in the last 12 months prior to the index calculation.
2. *Management Score for Milk (kg):* The general principle of the Management Score is to evaluate if the genetic potential of cows to produce milk in a herd is well expressed, isolating the "environment" portion of the popular "Phenotype = Genetics + Environment" equation.
3. *Transition Cow Index™:* Objective assessment of the success of the transition period at the herd level, based on Nordlund (2006).

Heifers

1. *Calf Mortality Rate (%):* Percentage of calves reported dead within the first 24 hours after birth over the last 12 months.
2. *Age at First Calving (months):* Average age at first calving of the animals that calved within the 12 months preceding the calculation of the SI.

Health

1. *Cows with BHB >0.20 mmol/ l milk (%):* percentage of cows with a result greater than 0.20 millimoles of Beta hydroxybutyrate (BHB) per liter of milk during the 12 months preceding the calculation of the SI.
2. *Cows >200,000 cells/ml (%):* percentage of cows that had a somatic cell count (SCC) greater than 200,000 in the 12 months prior to the calculation of the SI.

Sustainability Index Calculations

A percentile rank (1-99) is assigned for each of the ten parameters for each herd and the ten percentile ranks are summed to establish an overall ranking of the herds. This overall ranking is then re-expressed as a percentile rank, which is the SI of the herd. Herds can therefore compare against peers for their overall SI, but also for each of the ten indicators separately to better identify and improve weaknesses.

In order to have a SI calculated, a single herd must have at least seven out of the ten indicators available at milk recording (some are optional). For any missing data, a percentile rank of 50 is assigned. The SI is calculated for all Canadian herds three times a year (April, August and December). Each reports presents actual results as well as historical performance of the herd over the last two years to monitor improvement.

Results and discussion

A subset of 2,608 eastern Canadian herds were used to compare performance of top herds according to the SI versus the top herds in terms of performance (Table 1).

Overall, with increasing herd SI, the herd performance improved for production (+8.9 kg/cow and day; top 10% versus bottom 10% herds for all reported results), herd longevity (–36% involuntary culling and –74% cow mortality), heifer management (–62%

calf mortality), reproduction (−36 days open and −37 days for calving interval), health parameters (−51% potential subclinical mastitis; −71% potential hyperketonemia).

Simulations were done using a fictional herd with a quota of 100 kg of butterfat/day (Canadian supply management system and milk prices), Overall herd profitability was increased (+26% milk value and +$93,810 in operation margin) for herds in the top 10% for SI.

A descriptive comparison was made to compare performance and demographics of the best herds for SI versus the best herds in terms of milk value (milk yield corrected for price of milk components; Top 10% for both groups; Table 2).

Top SI herds had slightly lower corrected milk (-2.9 kg/cow/d) but improved milk quality and higher longevity due mainly to reduced culling.

Finally, a descriptive study of association with farm management practices on 2,143 dairy farms collected via a questionnaire (Lactanet, 2021) revealed some important aspects of calf, heifer, dry cow, and milking management. In particular, improving the stall surface and housing comfort, colostrum management, ventilation, and udder and hoof health were key aspects to improve the herd SI.

Table 1. Comparison of demographics and performance of bottom, average and top herds for the sustainability Index (SI) based on 2,608 eastern Canadian herds.

	Bottom 10% SI	Average SI	Top 10% SI
Production			
Fat (%)	4.16	4.10	4.09
Protein (%)	3.22	3.18	3,18
SCC ('000 s.c./ml)	263	193	137
Milk (kg/year)	8,414	9,844	11,023
Corrected Milk (kg/day)	31.4	36	40.3
Demographics			
3rd lactation or + (%)	37	42	46
Culled Cows (%)	39	34	32
Involuntary Culling (%)	23	19	15
Age at First Calving (months)	27.1	25.2	23.9
Reproduction			
Calving Interval (days)	432	409	395
Days Open	152	129	116

Table 2. Comparison of demographics and performance of the best 10% herds for sustainability index and milk value (CAN$/cow and day).

	Sustainability index	Milk value
Production		
Fat (%)	4.09	4.03
Protein (%)	3.18	3.15
SCC ('000 s.c./ml)	137	170
Milk (kg/year)	11023	11990
Corrected Milk (kg/day)	40.3	43.2
Demographics		
3rd lactation or + (%)	46	41
Culled Cows (%)	32	38
Involuntary Culling (%)	15	21
Age at First Calving (months)	23.9	24.2
Reproduction		
Calving Interval (days)	395	401
Days Open	116	121

Customized reports are published three times per year and made available to dairy producers and advisor services, including historical records to monitor changes over the last two years (an example as well as additional information on the SI can be found on online: https://lactanet.ca/en/thematique/farm-sustainability/).

Conclusion

In conclusion, milk recording data routinely collected and available in DHI databases was used to develop a herd sustainability index and compare herds across Canada. The SI can be useful to assess and monitor herd sustainability and can help dairy producers adopt best management practices to improve overall production performance and herd profitability.

Acknowledgement

Authors wish to thank the committee members who participated in the development of this index, as well as Lactanet colleagues and producers who provided feedback throughout development.

References

Lactanet. 2021. Portrait des fermes laitières Québécoises en matière de logement et de gestion des animaux. Lactanet, Sainte-Anne-de-Bellevue, QC, Canada. https://lactanet.ca/wp-content/uploads/2022/01/brochure_portrait-ferme-qc_021v11.pdf.

Nordlund, K. 2006. Transition Cow Index™. 39th Proceedings American Association Bovine Practitioners. St-Paul, MN, Sept 20-24. P 139-143.

Ryan, M., T. Hennessy, C. Buckley, E.J. Dillon, T. Donnellan, K. Hanrahan and B. Moran. Developing farm-level sustainability indicators for Ireland suing the Teagasc National Farm Survey. Irish Journal of Agricultural and Food Research 55 (2): 112-125. doi: 10.1515/ijafr-2016-0011.

Warner D., E. Vasseur, M. Villettaz Robichaud, S. Adam, D. Pellerin, D.M. Lefebvre and R. Lacroix. 2020. Development of a benchmarking tool for dairy herd management using routinely collected herd records. Animals 10 (9): 1689. doi: 10.3390/ani10091689.

Towards new breeding tools in a context of climate change: first results of the RUMIGEN project on new phenotypes for heat tolerance traits

S. Mattalia[1], A. Vinet[2], J. Vandenplas[3], M.J. Carabaño[4], C. Diaz[4], M. Ramon[5], S. Aguerre[1], B.C.D. Cuyabano[2], D. Boichard[2], E. Pailhoux[2] and H.A. Mulder[3]

[1] Institut de l'Elevage, UMT eBIS, UMR1313 GABI,
Domaine de Vilvert, 78350 Jouy-en-Josas, France
Corresponding Author: Sophie.mattalia@idele.fr
[2]Université Paris Saclay, INRAE, AgroParisTech, UMR1313 GABI,
Domaine de Vilvert, 78350 Jouy-en-Josas, France
[3]Wageningen University and Research, Animal Breeding and Genomics,
PO Box 338, 6700 AH Wageningen, The Netherlands
[4]Departamento de Mejora Genética Animal, INIA, 28040 Madrid, Spain
[5]Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal de
Castilla-La Mancha (IRIAF-JCCM), 13300 Valdepeñas, Spain

Abstract

RUMIGEN is a project financially supported by the EU that aims to develop breeding programs capable of managing the trade-offs between efficient production and resilience to extreme climate conditions. RUMIGEN is designed under a multi-disciplinary approach that mixes competencies in both genetics and social sciences. The genetic approach aims to enhance genomic selection using three levers: quantitative genetics, genome editing, and epigenetics.

One of the objectives of RUMIGEN is to enlarge selection criteria and to provide genomic tools to select dairy cows tolerant to heat stress. Studies are dedicated to the definition of heat-tolerance traits based on production, reproduction and health records, as well as to the study of the trade-offs between these traits, and with those already selected. These analyses are based on performances recorded in commercial herds in France, Spain and the Netherlands (i.e., milk production traits and somatic cell scores recorded by Milk Recording Organizations, and fertility traits derived from AI events), in combination with meteorological data obtained from the corresponding Meteorology Agencies. Records are associated to meteorological information at the farm level, in order to measure the impact of heat stress. First results obtained for different breeds and in a large range of farming and climatic scenarios showed that the combination of both types of information is relevant to measure the decline of performances due to heat stress and to define heat stress indicators and new traits for future breeding tools.

Keywords: Dairy cattle, animal breeding, heat tolerance, climate change, new phenotypes.

Introduction

Social expectations are more and more pushing for the development of sustainable breeding programs and linked technologies, as well as of breeding for adaptation to climate change while taking into account genetic diversity. The RUMIGEN program (https://rumigen.eu/), financially supported by the EU, aims to produce robust and efficient cattle able to manage the trade-offs between production and adaptation to extreme climate conditions.

To reach this goal, RUMIGEN gathers a large panel of competences. Research in quantitative genetics are focused on the study of heat tolerance traits and of the impact of climate change on the trade-offs between traits. The partners also study new approaches to maintain genetic diversity using genome information and the potential use of epigenomics information to enhance genomic predictions. Another ambition of this program is to improve the understanding on how genetic and epigenetic processes shape the phenotypes. Several experiments will be conducted to determine the impact of environmental stressors (metabolic demands during the gestation, heat, immunity challenge and pathogens). Epigenotyping tools will be used on a large scale to study the variability of the epigenomic profiles, their transmission between generations, and their impact on phenotypes. Regarding new biotechnologies, genome editing has been also suggested as a promising tool to speed up selection and design disease resistance alleles. However, it raises technical and social issues since only few data are available regarding the genome integrity (off-targeting, rates of de novo mutation). This is why RUMIGEN aims to assess the potential of genome editing to speed up introgression and the security of genome editing in different conditions. Finally, animal breeding brings not only biological, technical and economical issues, but also societal concerns with respect to breeding objectives, biodiversity and the potential use of new breeding technologies. Therefore, a work-package of RUMIGEN is dedicated to the social acceptance of breeding and the related technologies.

One of the main objectives of RUMIGEN is to provide is to provide breeding tools to face selection under the harsh environmental condititidions generated by climate change. For this purpose, France (Idele, INRAE), Spain (INIA—CSIC, IRIAF) and the Netherlands (Wageningen University and Research) collaborate to define new traits related to heat tolerance. Large-scale data from commercial farms are combined with meteorological information in order to highlight mid-long term impact of heat stress on performances and to define new traits related to heat tolerance.

The objective of this paper is to present the first results of using performances of Spanish, French and Dutch dairy cows recorded by Milk Recording Organizations (MROs) and associated meteorological data to measure the impact of heat stress period on production and health traits.

Material and methods

Performances of the cows used in the analyses

Test-day (TD) records for milk yield (MY, kg/d), fat yield (FY, g/d), protein yield (PY, g/d), fat content (FC, %), and protein content (PC, %) for French, Spanish, and Dutch cows, were extracted from the respective genetic evaluations for production and health traits. In addition to these traits, TD records for somatic cell score (SCS; defined as SCS = 3 + log_2 (SCC/100,000), with SCC being somatic cell counts in cells/ml) were available for French and Dutch cows.

Three breeds were involved in the analyses: the Holstein population, studied in the three countries, and two regional breeds, that is Montbéliarde studied in France, and Meuse-Rhine-Yssel (MRY) studied in the Netherlands.

For all countries, the extracted datasets covered approximately a period of 10 years, starting from the 2010s. Different edits were applied across countries. For example, after editing, the analyses carried out cover out the period 2016 - 2020 for France, and the period 2010 to 2020 and 2021 for Spain and the Netherlands, respectively. All countries analysed the different lactations separately with univariate models. Common edits were applied by all three countries on parity, age of calving at different parities (age of calving between 23 and 42 months and 35 to 60 months for 1st and 2nd parities respectively) and extreme phenotypic values. In all countries, only cows with known parents were considered. Only Days in Milk (DIM) between 5 and 305 were retained in France and in the Netherlands, between 5 and 400 in Spain.

Briefly, the datasets with production and health traits (MY, PY, FY, PC, FC, and SCS) for Holstein included around 36 million test-day records associated with around 7 million French first- and second-parity Holstein cows, around 20 million TD records associated with around 1 million Spanish first- to third- parity cows, and around 7 million TD records associated with around 500 thousand Dutch first- and second-parity Holstein cows. These cows were distributed in around 45,000 French herds, around 4,600 Spanish herds, and around 1,500 Dutch herds.

Weather data

Meteorological data were provided by Météo-France (Safran database) for France, by the National Meteorological Agency (AEMET) for Spain, and were extracted from the Koninklijk Nederlands Meteorologisch Instituut (KNMI) website for the Netherlands. Weather records were available for the French territory in the form of a grid of 9,892 8x8km squares, for 1,993 Spanish weather stations, and for 34 Dutch weather stations. Each herd was connected to theses meteorological information through the (partial) ZIP code of the farm.

For each daily record measured in each weather station, a Temperature – Humidity Index (THI) was computed using the NRC (1971) formula:

$$THI = (1.8*T+32)-(0.55-0.0055*RH)*(1.8*T-26)$$

with T being the average daily temperature (degrees Celsius) and RH the average daily relative humidity.

Models

The phenotypes (MY, FY, PY, FC, PC and SCS) of each lactation were analysed in separate studies. THI was calculated as the average THI of 3 days (in Spain and in the Netherlands: 3 days before the record; in France: day of TD and 2 days before).

For each population, the effect of THI on phenotypic performances at the population level was estimated using the following model:

$$y = X\beta + Z_1 a + Z_2 p + e$$

where y, β, a, p and e are respectively the vectors of phenotypes, fixed effects, additive genetic random effects, random permanent environment effects and the random residuals respectively, and X, Z_1 and Z_2 are the incidence matrices for the listed effects.

The fixed effects were almost the same in all countries, excepted the gestation stage that was included in the French and in the Dutch model but not in the Spanish one. However their combination differ from one country to the other:

- France: Herd-Year, THI, DIM, gestation stage, month of calving and age at calving.
- The Netherlands: Herd-Year, THI, DIM, gestation stage, age at calving-year- season.
- Spain: Herd-Year-Season, THI, DIM and age.

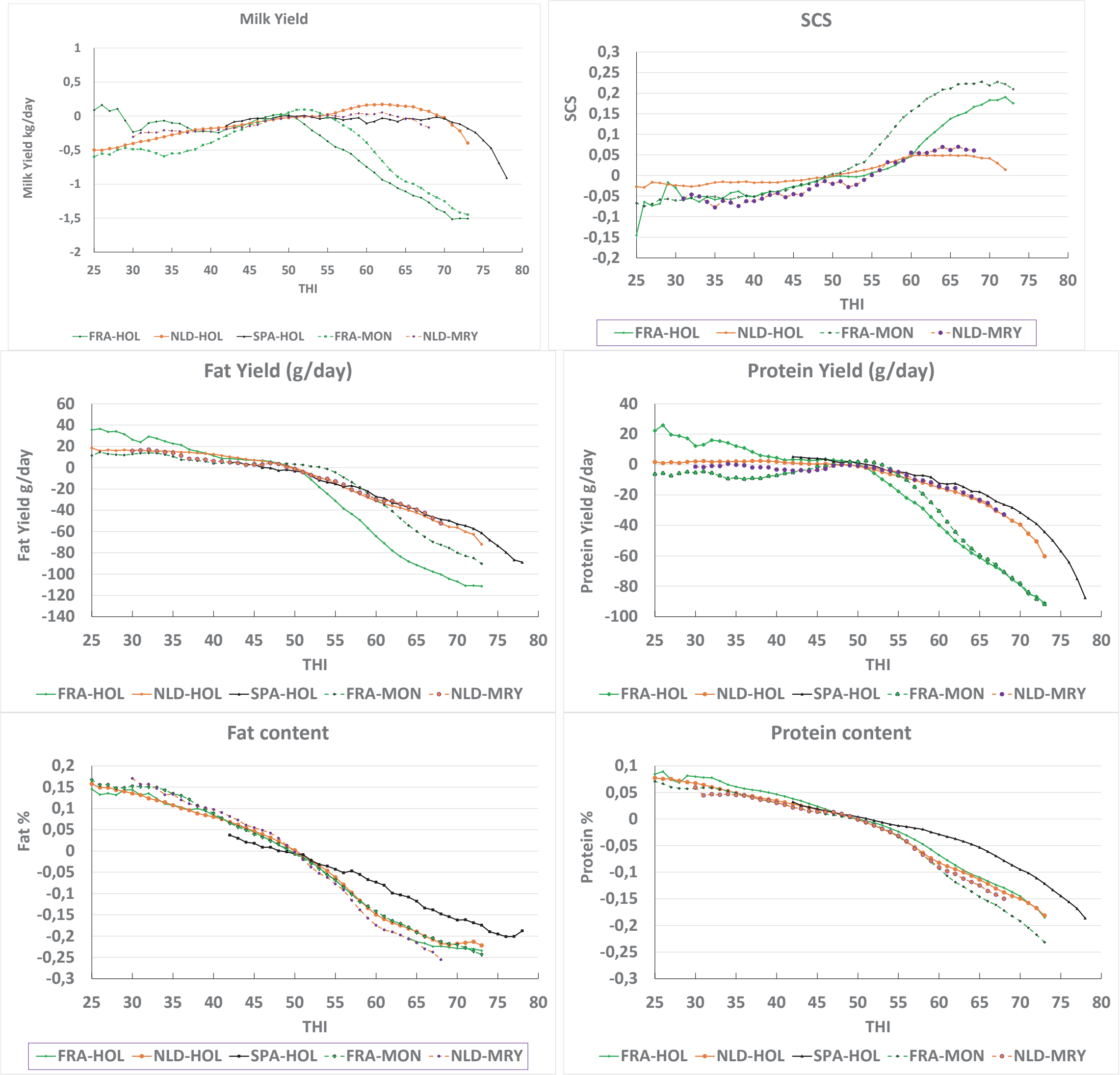


Figure 1. Effect of THI on MY, FY, PY FC, PC and SCS on cows in 1st lactation, according to the country and to the breed (FRA=France; NLD = the Netherlands; SPA = Spain; HOL=Holstein; MON=Montbéliarde, MRY=Meuse-Rhine-Yssel).

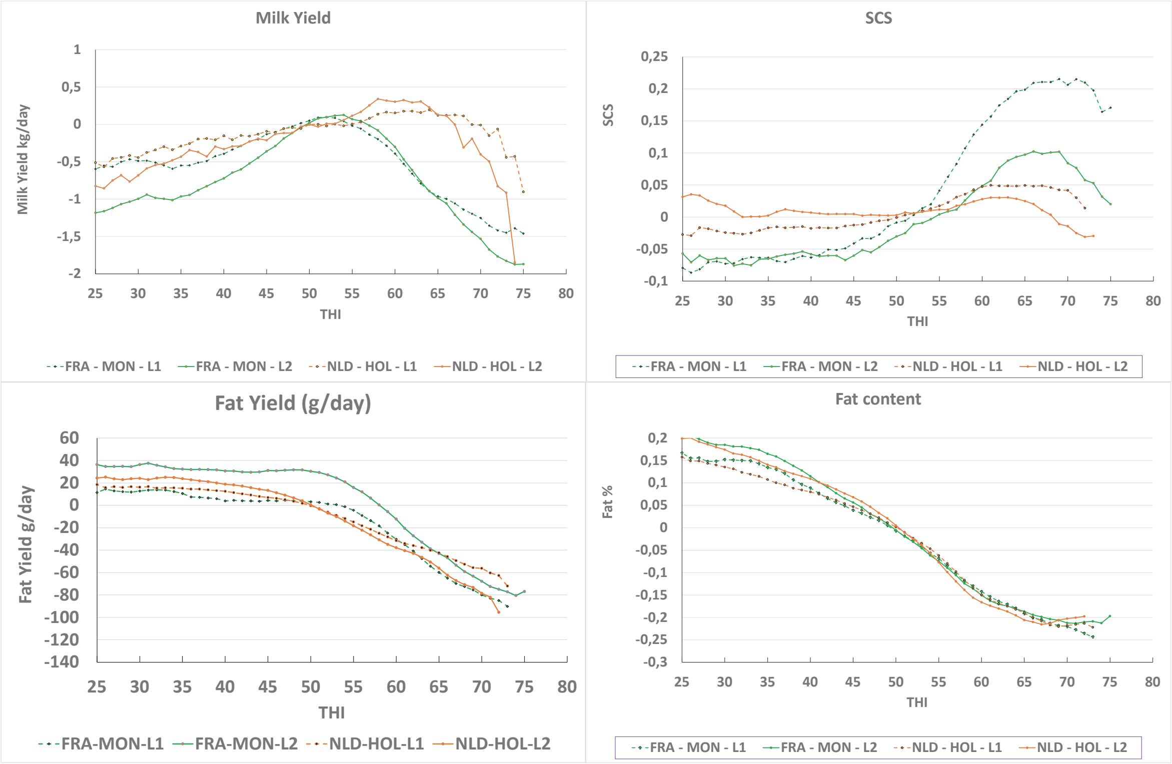


Figure 2. Effect of THI on MY, FY, FC and SCC on French Montbéliarde and Dutch Holstein cows in 1st and in 2nd lactation (FRA=France; NLD = the Netherlands; HOL=Holstein; MON=Montbéliarde).

Results and discussion

The estimated effects of THI on MY, FY, PY, FC, PC for 1st parity cows of the three breeds and of the three countries, and on SCS for Dutch and French populations are presented in figure 1. The estimated THI effects on MY, SCS, FY and FC in 1st and 2nd lactations are compared on figure 2 for French Montbéliarde and Dutch Holstein cows. Estimated THI effects presented on both figures were averaged using a rolling basis of 5 THI units to smooth the fluctuations.

This study covered several breeds, climatic conditions and farming systems. A common approach was defined between countries and breeds to identify heat stress indicators and to measure their impact on the cow performances.

Dutch results of estimated effects of high THI must be interpreted carefully since the Netherlands are exposed to oceanic climatic conditions, with moderate daily variations. Only 66 days with a daily THI over 72 were observed within 10 years (2010-2019). However, THI over 60 were observed on average 120 days each year during the same decade, which means that the effect of a moderate heat stress can be measured using Dutch data. The frequency of days with very high THI was much higher in France and in Spain (eg. in France, pending on the region, between 88 and 345 days with a THI over 72 observed within 5 years).

This study showed that increasing THI had a negative impact on all studied traits, but more prominent for production than for SCS. To illustrate that, the estimated decrease in FY between 50 and 70 units of THI (ie., 9 and 22 ºC of daily average temperature for the relative humidity in summer in France) corresponds to 11% of the daily production of 1st parity French Holstein cows. The pattern of the curves of THI effects depended on the trait: for MY, SCS and PY, the THI effects were almost stable below 50 (with some exceptions such as the French Holstein cows for PY), while a decline was observed all along the THI scale for FY, FC and PC.

Some differences of magnitude of the estimated effects and of THI-thresholds were observed between countries, more than between breeds, particularly for the yields and for SCS. For MY, FY and PY, moderate THI (50-70) had a stronger impact on the performances of French females than for the other populations. For SCS, the effect of increasing THI over 50 was stronger for the 1st parity French cows. The comfort regions observed in France were relatively low when compared to the literature (Carabaño *et al*., 2017) and to the two other countries. This could be explained by differences of farming systems, that are often based in France on pasture. Thus French cows are often exposed to outside temperatures (even if they are kept inside during very hot periods) which is almost not the case in Spain. In the Netherlands, a large proportion of females included in this study were raised in farms equipped with automatic milking systems, with less grazing and potentially equipped with barns more adapted to heat than the average Dutch farms. Brügemann *et al.* (2012) also reported lower milk yield losses in feeding systems based on crop production than on pasture. The exposure to wind, more frequent in summer in oceanic regions than in continental ones may also explain some differences. In Spain, a probable acclimatization along the large periods of high temperatures during the summer and the use of heat abatement devices and mitigation practices in many farms could also explain some differences with the two other countries.

Conclusion

This study showed that heat stress indicators could be defined, combining performance and meteorological data. Increasing THI, mostly due to increasing temperatures, had a negative impact on all studied traits and all breeds. The estimated impact differed from one trait to the other and it was more prominent for production than for SCS. Therefore, heat stress tolerance is a complex trait. The studies are going on with reproduction traits in order to get a global view on the major traits included in breeding goals in dairy cattle.

Differences of magnitude of the effects of heat stress and of THI-thresholds were observed between countries, more than between breeds. This could probably be due to differences in farming systems and, in the case of Spanish cows, an acclimatization along the summer periods.

Acknowledgment

This study received funding from the European Union’s Horizon 2020 research and innovation program under grant number 101000226 (Rumigen). This project adheres to EuroFAANG (https://eurofaang.eu). The CAICalor project was funded by APIS-GENE. The authors thank Meteo-France, AEMET and KNMI for the meteorological data and INRAE-CTIG, CONAFE and CRV for the performance and pedigree data.

The use of the high-performance cluster was made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR, Wageningen, the Netherlands).

References

Brügemann, K., E., Gernand, U. Konig von Borstel, and S. Konig, 2012. Defining and evaluating heat stress thresholds in different dairy cow production systems. Arch. Tierzucht 55:13–24.

Carabanõ M.J., M. Ramon, C. Diaz, A. Molina, M.D. Pérez-Guzmán and J.M. Serradilla, 2017. Breeding for resilience to heat stress effects in dairy ruminants. A comprehensive review. J. Anim. Sci. 95(4):1813-1826.

NRC. 1971. A Guide to Environmental Research on Animals. Natl. Acad. Sci., Washington, DC.

Practical experience from the implementation of a new udder health monitoring service through regular DHI testing

D. Schwarz[1], G. Witzel[2], S. Kleinhans[2] and K. Ilves[3]

[1]FOSS Analytical A/S, Nils Foss Alle 1, 3400, Hilleroed, Denmark
Corresponding Author: das@foss.dk
[2]Qnetics, Stotternheimer Strasse 19, 99087 Erfurt, Germany
[3]EPJ, Friedebert Tuglase 12, 50094 Tartu, Estonia

Abstract

This study aims to describe the practical experience of introducing a new udder health monitoring service in Estonia and Thuringia, Germany. This new service is based on the combination of somatic cell count (SCC) and differential SCC (DSCC) test day results obtained through regularly performed dairy herd improvement (DHI) services.

A new udder health report based on SCC and DSCC information was developed. It allows to categorised the udder health status of a cow into four different groups: Udder Health Group (UHG) A – healthy, SCC <200,000 cells/ml and DSCC $\leq$65%, B – onset of mastitis, SCC <200,000 cells/ml and DSCC >65%), C – (active) mastitis, SCC >200,000 cells/ml and DSCC >65%, D – chronic mastitis, SCC >200,000 cells/ml and DSCC $\leq$65%.

Regularly available DHI test results were used to investigate the performance and future development (e.g. high SCC at next test day, culling) of cows in the different UHG. Findings were used during the launch of the new udder health report as they contribute to provide evidence on the added value of the service based on local data. DHI service field staff was trained about the new service to support new herds enrolling to the new service. Different media and local events such as exhibitions and annual meetings were used to further promote.

Practitioners working with the new udder health report described that it helps them to improve their herd management, particularly cubicle management and milking routine, resulting in overall better udder health of the herd and lower consumption of antibiotics with regards to mastitis treatments.

Keywords: Mastitis, SCC, milk recording, antibiotics.

Introduction

The companies EPJ and Qnetics offer various services to dairy farmers in Estonia and Thuringia, Germany, respectively. Among other things, these include the organisation and execution of DHI testing for dairy cows, ewe, and dairy goats, milk payment analysis, and advisory services for herd management, dairy cow nutrition and milk quality challenges.

Dairy farming in Estonia and Thuringia, Germany

The total number of dairy cows is 80,000 and they are kept in approximately 400 herds, which translates into an average herd size of 200 cows. More than 96% of herds are enrolled to DHI services. The average annual production is at 10,484 kg milk, 394 kg fat, and 345 kg protein per cow.

There are 94,548 dairy cows on 282 dairy farms in Thuringia, Germany. The vast majority (98.2%) of these farms are utilising dairy herd improvement services on a regular basis. The average annual production is at 9,978 kg milk, 404 kg fat, and 345 kg protein per cow. The average herd size is at 335 cows and approximately 60% of all cows are kept on 25% of all farms.

The key motivation of EPJ and Qnetics was to be able to provide farmers with better DHI-based services for managing udder health, in particular earlier detection of udder health issues.

Udder Health Group concept

Regularly available DHI samples are analysed on CombiFoss 7 DC. The SCC and DSCC test day results are then used to were categorised the udder health status into four different udder health groups (UHG, Figure 1):

- A – healthy: SCC <200,000 cells/ml and DSCC ≤65%,
- B – onset of mastitis: SCC <200,000 cells/ml and DSCC >65%
- C – (subclinical) mastitis: SCC >200,000 cells/ml and DSCC >65%
- D – chronic mastitis: SCC >200,000 cells/ml and DSCC ≤65%

The SCC cut-off of 200,000 cells/ml used is according to IDF recommendations (IDF, 2013). The DSCC cut-off of 65% has been identified as optimal cut-off in a study where the actual udder health status of cows as defined by bacteriological testing was available (Schwarz *et al*. 2020a).

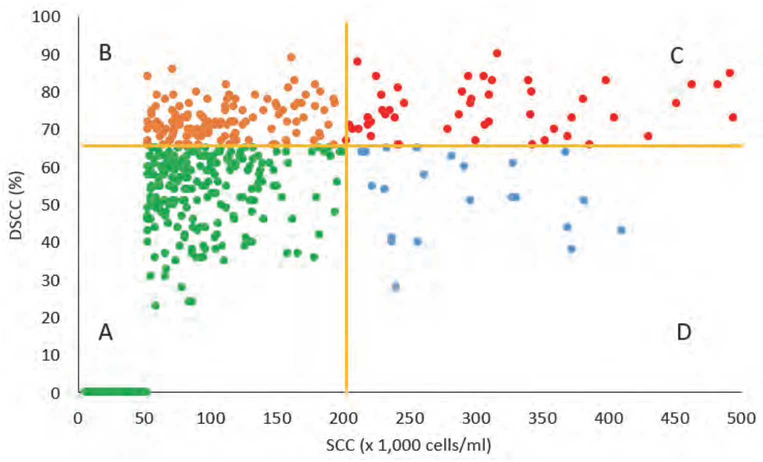


Figure 1. Dairy herd improvement results of one dairy farm (random example) to illustrate the four different udder health groups: A – healthy, B – onset of mastitis, C – (subclinical) mastitis, D – chronic mastitis.

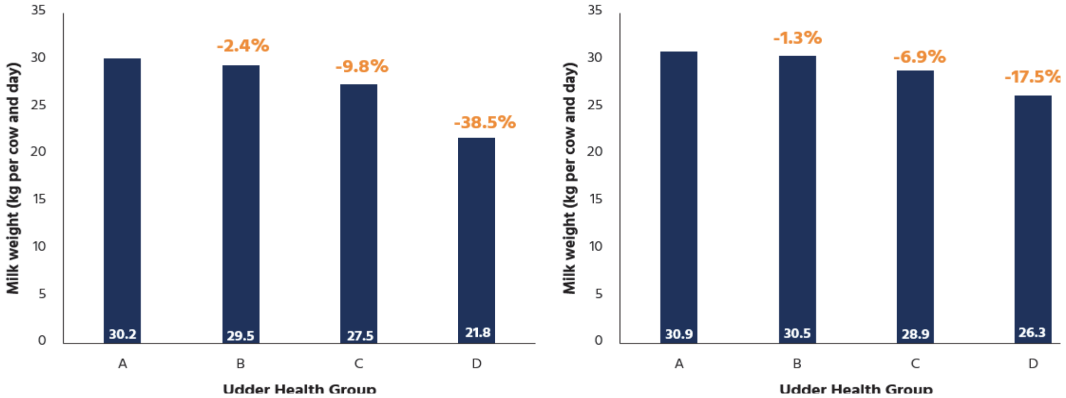


Figure 2. Average daily milk weights in different Udder Health Groups based on data from Estonia (left) and Thuringia, Germany (right).

Comparing milk weights in the different UHG (Figure 2) based on data from Estonia and Thuringia, Germany, interesting differences were found. These differences, in turn, indirectly confirm the interpretation of the four groups. Cows in group A showed the highest performance, which was to be expected as they are healthy based on their low SCC and low DSCC test day results. Cows in group B showed a significantly lower performance than those in group A, which is explainable based on inflammatory processes (i.e. high DSCC values) consuming energy that is then not available for milk production anymore. This confirms that cows in group B might be in the early stage of mastitis. Presences of mastitis pathogens and inflammatory processes has been described in such cows before (Schwarz *et al*., 2011a,b; Pilla *et al*., 2012; Schwarz *et al*., 2020a). Lower performance of cows with elevated SCC have been described before, but we could observe significant differences between cows in groups C and D. Those in group D performed significantly worse than cows in group C. Cows in group D are considered to experience chronic mastitis and it is well-know that such chronic infections lead to destruction of milk producing tissue explaining the low performance. All results have been published in detail elsewhere (Schwarz *et al*., 2020b).

Above-described results have been presented to dairy farmers at various events and contributed to create a lot of interest because most farmers would be interested in optimising the performance of their herds and understand the distribution of their cows among the 4 UHG.

New udder health report introduced

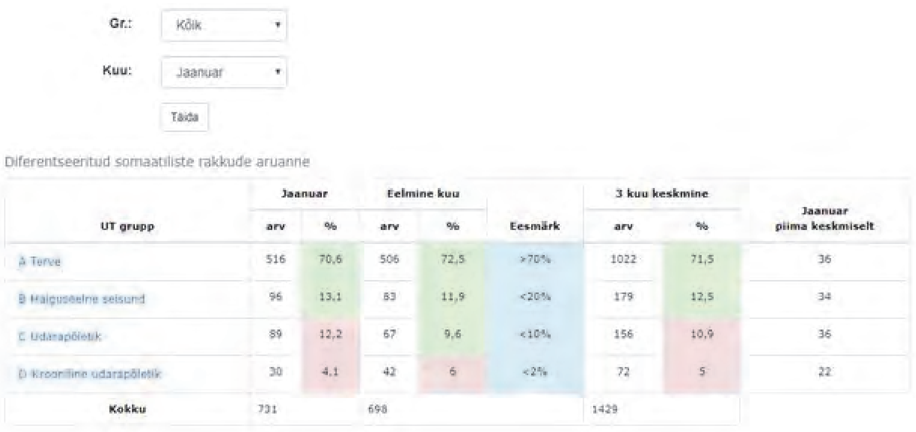
New udder health reports to visualize and process the combined results of SCC and DSCC were implemented in Estonia and Germany, respectively. Besides a graphical overview of test day results, a benchmarking table (Figure 3) is core of this report. The table provides a detailed overview on number and percentage of cows in each of the four UHG, target values, and average results of all herds tested. Percentage numbers are colour-coded to simplify the interpretation of the results.

Experience from the field

Herds working with the new udder health report agree that “the information provided through the new report and the arising possibilities aid in improving animal health and, at the same time, lead to better usage of resources and higher animal performance.” It allowed them to, among other things, fine-tune their cubicle management and milking hygiene and helped them to document improvements. This, in turn, was invaluable because staff could be motivated to do some extra effort and apply known measures used to prevent mastitis. Often in collaboration with veterinarians, herds also based selective dry cow therapy decisions on the information provided through the new udder health report. As a result, the number of antibiotic treatments and the volume of wasted milk were reduced.

In EU, a new Animal Health Law (http://data.europa.eu/eli/reg/2016/429/2021-04-21) has been introduced recently and is in force now. The livestock industry is supposed to avoid unnecessary use of antimicrobial treatments and, instead, focus on prevention of diseases. In this context, veterinarians appreciate the new udder health report because it helps them to convince herd managers to improve in terms of prevention of mastitis.

A comparison of herds with different proportions of cows in UHG A has been conducted (Schwarz *et al*., 2022). A key finding was that average daily production was significantly higher and bulk tank somatic cell count and number of antibiotic mastitis treatments were significantly lower in herds with high compared to herds with low proportions of cows in UHG A.



Gr.: Kõik

Kuu: Jaanuar

Diferentseeritud somaatiliste rakkude aruanne

UT grupp	Jaanuar		Eelmine kuu		Eesmärk	3 kuu keskmine		Jaanuar piima keskmiselt
	arv	%	arv	%		arv	%	
A Terve	516	70,6	506	72,5	>70%	1022	71,5	36
B Haiguseelne seisund	96	13,1	83	11,9	<20%	179	12,5	34
C Udarapõletik	89	12,2	67	9,6	<10%	156	10,9	36
D Krooniline udarapõletik	30	4,1	42	6	<2%	72	5	22
Kokku	731		698			1429		

Figure 3. Screenshot of benchmarking table from new udder health report of a randomly selected herd as introduced in Estonia.

Conclusion

Mastitis is still a huge challenge on dairy farms and is one of the key reasons for premature culling of dairy cows. The new SCC and DSCC-based udder health report helps dairy herds to better manage udder health as it provides further insights in herd management. We have already observed improvements in terms of dairy cow health, welfare, performance, and a reduced number of antibiotic mastitis treatments. It is expected that the percentage of herds utilising the new service will continue to increase and that longevity and milk quality will improve in the long run as well.

References

IDF, International Dairy Federation, 2013. Guidelines for the use and interpretation of bovine milk somatic cell count. Bull. IDF. 466/2013.

Pilla, R., D. Schwarz, S. Konig and R. Piccinini. 2012. Microscopic differential cell counting to identify inflammatory reactions in dairy cow quarter milk samples. J. Dairy Sci. 95:4410–4420. https://doi.org/10.3168/jds.2012-5331.

Schwarz, D., U.S. Diesterbeck, K. Failing, S. Konig, K. Brugemann, M. Zschock, W. Wolter and C.-P. Czerny. 2011a. Flow cytometric differential cell counts in milk for the evaluation of inflammatory reactions in clinically healthy and subclinically infected bovine mammary glands. J. Dairy Sci. 94:5033–5044. https://doi.org/10.3168/jds.2011-4348.

Schwarz, D., U.S. Diesterbeck, K. Failing, S. Konig, K. Brugemann, M. Zschock, W. Wolter and C.-P. Czerny. 2011b. Microscopic differential cell counts in milk for the evaluation of inflammatory reactions in clinically healthy and subclinically infected bovine mammary glands. J. Dairy Res. 78:448–455. https://doi.org/10.1017/S0022029911000574.

Schwarz, D., D.E. Santschi, J. Durocher and D.M. Lefebvre. 2020a. Evaluation of the new differential somatic cell count parameter as a rapid and inexpensive supplementary tool for udder health management through regular milk recording. Prev. Vet. Med. 181, 105079. https://doi.org/10.1016/j.prevetmed.2020.105079.

Schwarz, D., S. Kleinhans, G. Reimann, P. Stückler, F. Reith, K. Ilves, K. Pedastsaar, L. Yan, Z. Zhang, M. Valdivieso, M.L. Barreal and R. Fouz, 2020b. Investigation of dairy cow performance in different udder health groups defined based on a combination of somatic cell count and differential somatic cell count. Prev. Vet. Med. 183, 105123. https://doi.org/10.1016/j.prevetmed.2020.105123.

Schwarz, D., S. Kleinhans, G. Reimann, P. Stückler, F. Reith and S. Dano. 2022. Herd management practices associated with good, medium, and poor udder health as defined based on a new udder health group concept. (submitted)

ICAR member's needs regarding the use of robots and sensor data

C. van der Linde and M. Burke

ICAR, Utrecht, The Netherlands

Abstract

A survey was conducted among 13 ICAR full members (mainly milk recording organisations) about their needs and how they use robot and sensor data in their services. Main aims of the survey were to inventorise the purpose, availability and needs of the use of robot and sensor data and the role that ICAR can play for members in using these data for their services.

The survey was also conducted among 4 ICAR associate members (manufacturers of robots or sensor devices) about the value of ICAR certification and their needs for ICAR approval of robot and sensor devices.

All full members that participated in the survey are using daily milk yields from robots for at least the calculation of the milk yield at the day of milk recording. Few other organisations use some robot measured traits like milking speed and teat conformation traits in their genetic evaluation. Other robot or sensor data is hardly used for official milk recording or genetic evaluation, but are used for heat or health alerts with some robot data reported back to the farmer for management purposes.

For members, the main traits of interest from robot and sensor data are the standard milk recording components, (for management and genetic evaluation), as well as body weight, milking time and milking speed.

Not being able to use robot and sensor data in their services is a hurdle members would like to overcome. Members expressed the need for harmonisation of robot and sensor outputs, to help with the uncertainty about the quality of data (e.g. is calibration, maintenance done). Many members are in the process of implementing ICAR Animal Data Exchange standards to solve issues with harmonisation of output.

For data from sensors, like activity meters, there is the question of which modifications to the data needs to be done and how to incorporate that information into genetic models or services.

The need for ICAR certification of devices, especially for official milk recording was seen as a key, as also it is to maintain data quality in general. If devices would not meet the standards for certification, there is still value in ICAR assessing and validating the accuracy level of the data from such devices. The ICAR validation of devices or systems should be clear for all stakeholders (manufacturers, members and users).

For manufacturers the value of ICAR certification is high. It enables them to get access to certain markets, it enables their customers to use data generated by the devices in milk recording or genetic evaluation. It also proofs the ability of the device to generate quality data.

Regarding the value of ICAR approval of sensor or robot devices, manufacturers ask the question for who is the benefit of ICAR approval. And what the value is for farmers.

There needs to be a balance between costs and benefits for them. Another remark was that sensor devices and certain traits measured by robots are generally designed for herd management, the development of these devices was not targeted on ICAR certification. Manufacturers see the risk that approval of lower accurate devices would dilute the value of current ICAR certification.

Introduction

ICAR has a long history in providing certification of devices and guidelines for the recording of quality animal data. For official milk recording, data is collected with ICAR certified milk meters and samplers. Nowadays, the recording of animal data is rapidly expanding. Farmers have invested in robot or sensor devices and are often asking ICAR members to use these data for services like milk recording. But getting access to these data and being able to include these data in existing or new services can be challenging. Therefore, ICAR conducted a survey among 13 ICAR full members (mainly milk recording organisations) and 4 ICAR associate members (manufacturers of robots or sensor devices). Aims of the survey were to inventorise members needs of the use of robot and sensor data, the value of ICAR certification or approval for manufacturers and members and manufacturers needs regarding ICAR certification or approval of devices.

Material and methods

The list of questions for this survey were composed by the ICAR staff. The survey for ICAR full members consisted of 9 questions (see Appendix 1). Questions asked to the members were about the services they offer to their clients, the current use of and needs for use of robot and sensor data in their services, the main hurdles they face in not being able to use these data and their needs for certification or validation of robot and sensor devices by ICAR.

The survey for manufacturers consisted of 5 questions (see Appendix 2). Questions asked to the manufacturers were about the devices or products they sell to their customers, their need for and the value for them of ICAR certification or approval.

The questions were mainly open questions. The survey was conducted during March through May 2022, by video meetings asking the questions to the participants. There were 16 full ICAR member organisations approached, involved in at least milk recording, and 4 ICAR associate members.

Results and discussion

A total of 13 full members from 11 different countries participated in the survey. These participants were from Australia, Canada, Czech Republic, Denmark, Finland, France, Germany, Ireland, The Netherlands, Norway and the United States. All the participants are involved in processing milk recording data. Most of the organisations also carry out related services like animal recording, milk analysis in the lab and genetic evaluations.

All participants collect milk yields from robot devices. Some of them also collect other data like fat and protein percentage, somatic cell count, milking speed and teat coordinates. As main reasons for not using more traits than milk yields are mentioned a lack of standardisation of the output of robot devices across manufacturers, a lack of ICAR certification and problems with linkage between observation and official ID. Most participants use the milk yield data only to calculate the milk yield at the day of milk recording, some use daily milk yields also to calculate lactation yields. But other traits than milk yield, measured by robots or sensors, are not used for other official services like milk recording, genetic evaluation or pedigree certificates.

Some members are investigating opportunities to include milking speed data from robots in the genetic evaluation of milking speed. Daily yields, heat alerts and other sensor alerts or data are used by some members for management reporting. Many members are considering to implement the ICAR ADE/iDDEN standards, some of them already have started implementing.

The data from robot or sensor devices that members would like to use for their services are components (fat and protein), milking time, milking speed, box time, body weight, body condition, activity data, rumination data. For components that are measured by robots (or any other on-line or in-line device) two conditions are mentioned though, the accuracy of the components measured needs to be known and there needs to be a frequent calibration of the device. Many members would like to use these components data for the milk recording and some also indicated to want to use it for the genetic evaluation.

Some of the most frequent hurdles mentioned that prevent the use of robot and sensor data for services to dairy farmers include; access to data, data collection (in regard of demand of time for milk recording), lack of advantage for the farmer to share data, lack of harmonisation of outputs, getting the same information from different systems, quality of data (accuracy, consistency, calibration and maintenance of device), source of information (version of software and hardware) not always available, missing certification of data and output and missing information about definition of data value and missing documentation.

Needs of ICAR members regarding certification of sensor or robot devices

Members clearly state that ICAR certification is important to maintain the quality of data at the current level. The quality of data is especially important to them for use in the genetic evaluation. Members are also interested in ICAR approval of devices that record data that does not meet criteria for ICAR certification, provided that:

1. They know the accuracy of the data measured.
2. They know with which type of device the data is measured.
3. ICAR provides recommendations about the required quality of data for certain purpose (e.g. genetic evaluation, milk recording, herd management etc.)
4. ICAR provides guidelines or recommendations about how members could use these data for different purposes.
5. Manufacturers provide information about the quality of the recorded data, about the relationship between the measured and real data value, about the proper use, maintenance and calibration of the device etc.
6. Data output across manufacturers of data measured on the same traits will be standardized as much as possible.

An important point of discussion is how ICAR could approve devices that do not meet current certification criteria. An even more important question is probably what kind of approval to assign to a potential ICAR test for these devices. The outcome of the test should be valuable for the manufacturer and the costs of the test should be balanced with the value for them. ICAR and ICAR test centres should be able to come up with clear and consistent test plans for these devices.

There were 4 manufacturers participating in the survey. All of the manufacturers are producing ICAR certified devices for milk recording, some of them are also producing other recording devices like sensors.

For manufacturers the value of ICAR certification is high. It enables them to get access to certain markets, it enables their customers to use data generated by the devices in milk recording or genetic evaluation. It also proofs the ability of the device to generate quality data.

Regarding the value of ICAR approval of sensor or robot devices, manufacturers ask the question for who is the benefit of ICAR approval. And what the value is for farmers. There needs to be a balance between costs and benefits for them. Another remark was that sensor devices and certain traits measured by robots are generally designed for herd management, the development of these devices was not targeted on ICAR certification. Manufacturers see the risk that approval of lower accurate devices would dilute the value of current ICAR certification.

Summary and conclusion

Daily milk yields from robots are widely used by ICAR members to calculate milk yield at the day of milk recording. Robot measured traits like milking speed and teat conformation traits are used by a few members in their genetic evaluation. Other robot or sensor data is hardly used for official milk recording or genetic evaluation, but e.g. heat or health alerts or some robot data are reported back to the farmer for management purposes.

Main traits from robot and sensor data that members want to use for their services are the components for milk recording and for the genetic evaluation, body weight, milking time and milking speed.

The main hurdles preventing ICAR members from using robot and sensor data in their services are: 1) need for harmonisation of robot or sensor output, 2) uncertainty about the quality of data and 3) unknown which modifications to the data needs to be done to incorporate this information into genetic models or services. Many members are in the process of implementing ICAR Animal Data Exchange standards to solve issues with harmonisation of output.

ICAR certification or validation is important for ICAR members to maintain the quality of data at the current level, especially for milk recording. If data recorded by devices does not meet criteria for ICAR certification, members are still interested in ICAR approval of these devices to use these data. In that case it is important to know the accuracy of recorded data, the type of device that recorded the data and to get recommendations how and for which purpose to use the data.

Manufacturers play an important role in providing information about the quality of recorded data and about maintenance of the device and by standardising data output. The ICAR validation of devices or systems should be clear for all stakeholders (manufacturers, members and users).

ICAR certification enables manufacturers to get access to certain markets. Therefore, the value of ICAR certification is high for them. It also enables their customers to use data generated by the devices in milk recording or genetic evaluation. ICAR certification also proofs the ability of devices to generate quality data.

The main question of manufacturers about the value of ICAR approval of sensor or robot devices is who will benefit of it. What is the value for farmers, for manufacturers and for milk recording or related organisations? For manufacturers, there needs to be a balance between costs and benefits. Manufacturers also see a potential risk that approval of lower accurate devices would dilute the value of current ICAR certification.

Appendix 1. Questions of survey among ICAR full members about their needs for use of robot and sensor data.

Which services do you offer to your clients?

1. Animal identification
    - Milk recording (cattle)
    - Beef recording (cattle)
    - Herdbook recording
    - Conformation recording
    - Data processing
    - Laboratory analysis (milk)
    - Laboratory analysis (DNA)
    - Genetic evaluation (dairy cattle)
    - Others…
2. Do you make use of/include robot data in your milk recording service? If so, which data and how many herds/cows/records?
3. Do you flag/mark the data coming from sensors or robots to discern it from data sourced from approved milk recording sampling?
4. Do you restrict the use of sensor data? e.g. Ok for management but not for genetic evaluation / pedigree certs?
5. Do you make use of/include other robot or sensor data in other services? If so, which data for which service?
6. For which of these services do you want to make use of robot or sensor data and which data do you want to use?
7. What are the main hurdles to not be able to use robot and sensor data for your services?
8. What are your needs regarding certification of sensor or robot devices? What is the value for your organisation?
9. What do you want ICAR to develop regarding the use of robot and sensor data for your services?

Appendix 2. Questions of survey among manufacturers about the value of ICAR certification and their needs for ICAR approval of devices.

1. Which kind of devices do you produce/sell to your customers?
    - Milking robots
    - Sensor devices
    - Other recording devices
    - Data processing
    - Others…
2. Are there any of these devices you would like to be certified by ICAR?
3. What is the value for your company of ICAR certification?
4. Is there merit in ICAR certifying your sensor devices or specific traits measured by your robot?
5. Open discussion on the questions above.

Supporting automatic milking farms in milk recording

K. Leppikorpi

Association of ProAgria Centres, P.O. Box 251, 01301 Vantaa, Finland
Corresponding Author: kirsi.leppikorpi@proagria.fi

Abstract

In Finland there are 3755 herds in milk recording (73 % of all herds) and 200 134 cows (80 % of all cows). Currently, 52 % of the milk recording´s milk comes from robotic herds, and the estimated amount by 2035 is 80 %. More than 90 % of the Finnish farms are doing their milk recording work by themselves, such as milk sampling and data transferring, but there are several software and customer support services available. The calculation of milk recording data, reporting to customers and the maintenance of the programs are handled by Mtech Digital Solutions Ltd.

The Finnish milk recording organization ProAgria offers various services to all farms, but here the focus is on services for automatic milking farms (AMS) to help their daily work and keep them involved in milk recording. The basic services provided to AMS farms are different technical milk sampling services, milk recording and sampling data´s transferring service and customer service. In addition, AMS farms are offered a new comprehensive service that considers a whole farm. This service is still under development and is therefore called “AMS development program”.

Keywords: AMS, milk recording.

Introduction

Most of the Finnish AMS-farms do regularly milk recording and milk sampling by themselves without difficulty. However, part of farms finds this laborious and would like to either outsource the work or stop taking milk samples. There would be more demand for outsourcing these if more services were available.

There are delays in the data recording on some farms, which is reduced by offering a data transfer and recording services. The Data Quality Points system has helped reduce data recording delays and farms can monitor their officiality in milk recording monthly.

Services for robot farms

Data transfer and recording

Milk recording service is offered for robot farms. The expert will take and save the milk data remotely from the customer’s robot and checks at the same time all relevant information has been sent on time and the reports are fine..

Customer service

Customer service assists all the farmers in milk recording, but they also support robot farms with their many technical issues in milk sampling. At the same time, they monitor the situation of data quality points of the farms.

Robot sampling

There are plans to develop different new support services for robot sampling:

1. Continuous basic service, sampling 1 x / month by a technician.
2. Temporary service, according to the customer's needs, e.g., during seasonal work or in case of sickness etc.
3. Customer sampling support service; the customer takes care of the sampling, but can get help in many ways, if needed.
4. Help desk -phone, technical support.
5. Equipment rental (samplers, scanners).
6. Maintenance service for sampling equipment.

Some of these services already exists, but availability is currently limited to certain areas.

Challenges

Connection problems on farms causes the data transfer to be interrupted. There are only a limited number of technicians with technical expertise and located in certain areas; training and new employees are needed. Long distances between facilities and reaching the customers who benefit the most from the services are also challenges.

Conclusion

Services may not be needed immediately on a large scale because farms are used to doing it themselves and there is no ready-made model. If there is no visible order for the service, it is easy to think that there is no need for it. In Finland, the challenge of the robot sampling service is the long distances between the farms, it is difficult to make it profitable. In addition, other work should be arranged for the employee in connection with farm visits. However, the number of AMS farms using the data transfer and recording service has increased, as well as using milk recording customer service´s advices for milk sampling. So, we are going in the right direction.

Embracing new streams of dairy data - how good is good enough?

J. Mattison and S. Sievert

National DHIA/Quality Certification Services, P.O. Box 930399, 53593, Verona, WI, USA
Corresponding Authors: jmattison@dhia.org; sjsievert@dhia.org

Abstract

With the investment of new technologies for on-farm data measurement and recording, there is an increasing amount of data available from dairy operations. These data include estimates of traditional measures such as milk yield and components, but also provides access to measurements on new traits of interest to herd recording organizations. With the volume of data available, it is logical to find new paths for data movement from on-farm software or manufacturer servers into organization databases. ICAR has supported milk recording organizations and dairy producers in the traditional test-day collection of data with standards and equations that deliver data for daily management decisions and to qualified information for genetic evaluations. Today, the challenge that exists for recording organizations and national databases is balancing the volume of data from new sources with the accuracy of the individual measurements. Further, the speed of data flow has the potential to allow unqualified data to enter databases before issues or concerns are identified. The challenges of veracity, volume and velocity of new streams of data lead to the question 'how good is good enough' for usability by organizations and databases.

In addition to identification of the source, an evaluation of the data quality at a system level is key for determining the usability or suitability. Focus on the automatic identification as well as the automatic recording of measurements is critical, both on-farm and as data moves into central databases. In addition to linkage of animal ID to measurements, considerations on the completeness and consistency of continuous data flow are of merit. Finally, the editing of data points, including estimation of missing observations and the precision of individual measurements are a component of data quality and contribute to the usability of data. A fresh approach on the inclusion of new streams of data is needed. This systems quality approach should embrace the technology of on-farm sensors that generate valuable insights into dairy cattle milk yield, milk composition, conformation, behavior and activity with the goal of inclusion of usable data that is 'good enough' rather than focus on data exclusion based on precision of an individual measurement.

Background and perspective

Herd recording programs have traditionally relied on data collected on the dairy during periodic visits and includes milk yields and cow event data accompanied by visible cow identification by organizational staff. The source of the yield data, and subsequent milk component estimates, is ICAR-certified devices either owned by the herd recording organization or by the dairy operation. Even when properly installed and operated under ideal circumstances, the data provided is limited to the recording day and only provides an estimate of cow performance during that timeframe. This system has provided excellent data for both management systems and genetic evaluations as the accuracy of these measurements is essential for modelling the performance of cows across the interval between visits. However, the development of new technologies for measurement of various cow traits that include both traditional yield and component measures, but also data that physically describes cow conformation, movement, and even provides estimates of cow health and well-being. The adoption of these technologies by dairies has led to more data available to herd recording organizations and their customers cooperatively providing data with the expectation of real-time access to results and information for management decisions affecting all aspects of the dairy operation.

While there is, and will be, a segment of dairy operations participating in herd recording programs in the traditional fashion, there are opportunities for new streams of dairy data that may be used and useful to herd recording organizations. However the question that faces the industry with these data streams is how or can we use these data? The volume of data and the velocity of data generation is staggering relative to traditional data ingestion models. But with this data flow also comes an increased variation in measurements and questions on the validity of the data as it flows automatically from the dairy operation to external databases. In summary, how good is good enough? This paper does not propose to set standards or guidelines for use of data but does offer a review of considerations for herd recording organizations as they look to development of programs and services offered to the modern dairy operation.

Data use and quality

New streams of dairy data should be reviewed for the use, useful and usability of the data on various levels. Data used for management decisions or animal health/welfare metrics will most likely have different measures of quality associated with it compared to data used for genetic evaluations. However, this does not mean that data use is exclusionary for these purposes, rather there needs to be an evaluation of data sources and continuous monitoring of quality. One of the key considerations with new streams of dairy data is the dynamic nature of data collection/measurement. Many of these systems rely on the simultaneous recording of animal identification and measurement of the traits(s) by the system. While the location or pen identification may be more important for certain management traits relating to feeding or milking metrics, the linkage of animal identification used for data measurement with the official identification of the animal is an important part of the data quality. Herd recording organizations should have or will need to develop protocols for evaluating both on-farm and data transfer for identification linkage.

As previously noted, data quality needs to be monitored as we look to new streams of dairy data. Traditionally, the accuracy of the individual measurement has been considered critical to the assessment of usability and said accuracy is an important part of data quality. However, quality data has additional elements that must be considered as we look to use and usability in the volume of data available on many dairy operations. These quality elements include accuracy, completeness, consistency, credibility and custody of the data. A continuous approach to evaluation and monitoring of data quality is paramount for herd recording organizations. For some data measurements, bias and precision are important are varying levels, but other sources of error may negatively affect the usability of data for that metric or measurement.

While some of the quality elements can be identified and managed by assurance programs, there are elements that require external review by ICAR or other organizations to understand data handling in both farm management software and as the data moves to external databases. These include an understanding of data smoothing, estimation of missing measurements, and the distribution of errors relative to the collection system. External validation of these systems along with internal assurance programs is essential for determining usability however should not be viewed as exclusionary for data ingestion by herd recording organizations. Rather, consistently recording the source of the data measurement(s) in the data flow will allow for usability decisions to be made as the use of data is determined.

How good is good enough?

There is no simple answer to the question ‘how good is good enough’ in this presentation or in practice. However there is opportunity for new and dynamic approaches to evaluate new streams of dairy data by herd recording organizations. This approach should have five key considerations that include:

1. Validation of system data quality and recording the source of data measurements in the data flow.
2. Focus on the ingestion of data from farm management software instead of exclusion of data based on measurement accuracy.
3. Development of routine data quality monitoring tools to assure consistent and credible data.
4. Focus on the data use, usefulness, and usability for specific needs rather that an ‘all or nothing’ approach to data use.
5. Efficient removal of incomplete data instead of attempts to edit or standardize the data to meet arbitrary standards or improve the data quality.

This approach is a departure from traditional data collection and use schemes but is essential with new streams of dairy data. The volume and velocity of data may be viewed as a deterrent but with a data quality monitoring program, there is opportunity to deliver information to dairy operations, herd recording organizations, and national/multi-national databases.

DataHUB360°, data at the service of breeders

D. Saunier[1], J.B. Daviere[2], T. Decers[3], T. Aubry[4] and S. Desvois[5]

[1]Eliance, 149 rue de Bercy, 75595 Paris Cedex 12, France
Corresponding Author: david.saunier@eliance.fr
[2]Seenovia, 141 Bd des Loges, 53940 Saint-Berthevin, France
[3]Eaden, La Sablonnière, 72800 Le Lude, France
[4]Littoral Normand, 14 rue Alexander Fleming - 14204 Herouville saint Clair, France

Abstract

The evolution towards the automation of dairy farms is booming in the agricultural world, the digital data of breeders must take the turn of high-speed exchanges and allow valuation for the purposes of advice, calculation of indices for the genetics, but also predictive models on milk production and also on the management of their herd with short and long term modelling.

With the arrival of robots, connected milking parlours and sensors installed directly on the animals, breeders' demands are exploding. Each automaton needs individual data such as the inventory and the events of all the animals to ensure their operation and feed their algorithm to alert breeders on a daily basis in terms of heat detection and health.

Keywords: Cows, data exchange, platform, robot, sensor.

Introduction

The Eliance federation and its network have been tasked with developing a high-speed data exchange platform called dataHUB360°, to meet the demand of breeders and automaton manufacturers for the import of data in real time and the daily export. Today, each breeder in the process of installing an automaton requests the dataHUB360° link to have the history of inventory data and events for his breeding, and avoid entering this information manually.

The second step was to export the sensor data and ensure its daily delivery to the evaluation database.

The enhancement of this data has made it possible to offer new services to breeders, a comparison by type of feeding system, in connection with the milk production of his herd, but also to secure the data feeding the genetic system of tomorrow based on actual farm data.

Valuation of the data for advise the farmer

Function of datahub 360°

The dataHUB is a Web platform, dedicated to the dairy farmer which allows communication with APIs (Cloud manufacturer), but also with automatons on the farm, it is intended to be multi-platform and bi-directional in order to enhance the entry of data on the farm automatons, and avoid double entry by the breeder.

The main objective is to facilitate the exchange of information between the automatons deployed in the farms and the data present in the databases of the Livestock Consulting Companies.

To exchange data, the security of exchanges takes on its full extent, as a result the dataHUB contains a consent module and generates encryption keys specific to each farm, in order to guarantee perfect confidentiality of this personal data.

The system is also able to monitor data exchanges in order to know in real time the information that passes through the import and export of the breeders' automatons.

A monthly history will also be kept to know the data exchanged with the automatons of farms, in order to guarantee to the breeders a routing of the data on his breeding.

The system is able to connect to farm automatons installed on computers connected to the Internet network, but also on manufacturer Cloud systems.

All exported data is stored in Eliance Livestock Companies database and is used to feed applications for official milk performance as well as other milk performance tools focused on milk meter monitoring but also many other applications as indicated in following

Use the data for genetic evaluation

The datahub solution makes it possible to export the production data of each dairy cow which is stored on the milking parlour automatons equipped with milk meters and identification, during the milk performance control carried out by Eliance Livestock Companies.

These data are processed and then formalized by Eliance Livestock Companies before being sent to the central genetic database.

OriwWeb, tools for robot

OriWeb was designed to meet 3 main objectives:

1. Ensure the processing of daily robot data in order to be able to introduce them into technical and genetic information systems such as 24-hour checks
2. Optimize robot data processing processes to facilitate and automate the marriage of robot and laboratory files during the milk performance.
3. Allow the breeder to have these results quickly as soon as they are processed by email.

The calculation of cows' milk production will now be carried out, as recommended by ICAR, over 96 hours instead of 48 hours, taking into account all production.

This change makes it possible to align with international standards, to get closer to manufacturers' averages and to reduce the effects of performance control on production in the context of busy stalls.

Several levers have been activated to gradually move towards full automation of the robot performance check processing chain.

An intelligent engine will search for the robot file information in the data collected by dataHUB, eliminating the need for manual robot file collection and integration operations.

Centralization of the calculation engine which processes the data as soon as they are made available by the dataHUB platform.

OriWeb has been designed to adapt to the specificities of each company, by offering a very wide level of configuration, which can personalize, restrict or not use the modules offered.

ADesse: tools for monitoring Milk Meter

A follow-up module by monitoring milk meters allowing both early detection of meter malfunctions and extending the validity of the checks beyond 12 months, as long as the monitoring checks confirm the validity of the installation.

This new module, currently intended for checks on milking parlor with milk meter, is a real source of savings for Eliance Livestock Companies who can reduce the frequency of milk meter check visits. This method uses a Dynamic Linear Model (DLM, West and Harrison, 1989).

The average milk yield per stand and milking session is calculated over all milking on that stand. The resulting stand average is compared with the overall average. The deviation will be close to zero for a properly working meter. A DLM is based on a comparison per milking session of the average per stand with the overall average

Res-Previ: tools for dairy forecast

Milk forecasting is an important issue for dairy cooperatives or milk producer organizations.

Res'Prévi is the tool for producing and promoting milk production predictions for producer organizations and dairy cooperatives.

Res'Prévi uses two types of data: Data directly from dairies and data from milk performance. The dairies provide monthly milk delivery data for each farm as well as ancillary data (contracted volume and average milk price in the area). Eliance Livestock Companies provide all milk performance data, with also the data from dataHUB , animal variations in production as well as feed data.

The Machine Learning algorithm that was produced by the DataLab uses all the data described above to observe the impact of the variation of each of them on future milk production.

The model therefore learns from history to predict future data. The characteristic of Res'Prévi is that the model receives data very regularly, which allows it to adjust the prediction week after week.

The data from the dairies are transmitted each month and integrated by the DataLab.

Today this service is offered to 12 producer organizations and 1 dairy cooperative, representing 2,000 farms in France.

GenoCells®: tools for heath detection

GenoCells® is a revolutionary technology to determine with a high accuracy the SCC of each cow directly from the DNA analysis of a tank milk sample. Used in France since 2018, this technology is based on the correspondence between animal genotypes (= genetic identity) and presence of their DNA via their somatic cells in the mixing milk sample. The SCC results from this disruptive genomic method are as accurate as traditional flux cytometry method (r^2=0.99).

GenoCells® is more practical than a classic milk control operation because only one tank milk sample is necessary. This method can be performed several once in a year and is less expensive by 20% compared to the classic method.

The dataHUB platform makes it possible to retrieve the milk production of each cow between 2 milk tank pick-ups and allows this tool to save significant time for the return of the results to the breeder, whether for robots or milking parlours automated.

RoboMat: tools for robot valuation

RoboMat is a data valuation software from robots and automated milking parlors of all brands.

All data extracted by dataHUB is used by the Robomat application. The goal is to provide appropriate and specialized advice to breeders equipped with automatons,

The system analyses herd performance quickly with or without performance monitoring.

The system produces indicators with real added value and complementary to those used in manufacturers' software. Data from dataHUB is routinely updated, every night at 5am.

The system makes it possible to define production objectives at the farm and to compare with other farms that are in the same production system. (Notion of benchmarking)

The use of filters (stages of lactation, type of livestock, type of cow, etc.) makes it possible to finely analyse the animals in the herd and see which ones are less efficient.

The system also tracks production level and cows that are less productive

Medria Solutions: sensor for cow monitoring

Medria Solution supports breeders, advisory bodies and sectors by distributing monitoring tools and enhancing the resulting data.

The dataHUB platform is connected directly to the Medria API in order to import all farm data, and to export alarms in real time to other applications to carry out valuation and consolidate them with dairy data to alert the breeder.

Medria sensors monitor heat, calving, rumination and also animal comfort.

Work is in progress on thermal stress in order to measure the impact on milk production and to make farmers aware of the importance of animal welfare.

Research: Influence of liveweight kinetics for insemination.

A study was carried out on the kinetics of the animal's weight during breeding. All data was obtained from the dataHUB platform and made it possible to perceive the importance of the energy balance during breeding, other studies are needed to confirm the influence of weight on breeding.

Conclusion

Today, the data are in the farms and it is the farmers equipped with milking robots, automated milking parlours or sensors who are the data providers.

The data begins to circulate for the needs of the breeders first to feed their automaton with inventory and event data. The use of data from automated breeders is developing, but the breeder must give his consent to use them and define the scope of their use. Data multiplexing is the next step because monitoring sensors or robot software also need other data from other controllers or sensors to operate.

The valuation of data in breeding is only beginning and the automatons must increase the scope of exportable data in order to allow the breeder to be able to share more data with his advisers, his experts in order to guide him, and advise him on the management of his herd.

Health data is an important issue, and must be harmonized to be usable for advice and study purposes.

New perspectives on artificial intelligence are in progress, but the algorithms need a lot of data to be robust, and research work is necessary for an optimum valuation of all this data.

Magnitude of the reduction in automatically scored body condition from calving to nadir body condition score affects the fertility of Holstein cows

P. Pinedo[1], D. Manríquez[1,2], J. Azocar[3], A. De Vries[4] and I. Klaas[5]

[1]Department of Animal Sciences, Colorado State University, Fort Collins, CO, USA
Corresponding Author: pinedop@colostate.edu
[2]Department of Veterinary Economics, National School of Veterinary Medicine, Toulouse, France
[3]DeLaval Inc, Bannockburn, IL, USA
[4]Department of Animal Sciences, University of Florida, Gainesville, FL, USA
[5]DeLaval International AB, Tumba, Sweden

Abstract

Automatically generated body condition scores (BCS) through image technology enable daily assessments of body energy reserves of dairy cows. The availability of high frequency data allows for the analysis of specific patterns or points of interest, such as nadir BCS, and could result in quick interventions if necessary or tailored management adjustments. The objective of this study was to evaluate the effect of the decrease in BCS from calving to nadir BCS before first artificial insemination (AI1) on pregnancy per AI1 (PAI1) in Holstein cows. A retrospective observational study was completed using data collected from 6,100 lactations (primiparous = 3,683; multiparous = 2,417) starting between April 2019 and March 2021 in a commercial dairy operation located in Colorado, USA. Scores generated by BCS cameras (DeLaval International AB, Tumba, Sweden) at calving (BCScalv) and nadir (BCSnadir) were selected to calculate the ratio BCSnadir/BCScalv (BCSratio).

The BCSratio is a representation of the BCS change from calving to nadir, where greater BCS loss results in smaller values for the ratio. To facilitate the calculation of PAI1 probabilities, the resulting BCSratio values were categorized as low (≤ lower quartile, large BCS decreases), medium (interquartile range, moderate BCS decreases), and high (≥ upper quartile, small BCS decreases). Data were examined using logistic regression by univariable models that were followed by multivariable models considering calving season, occurrence of disease, and milk yield up to 60 DIM as covariables. All the analyses were performed separately for primiparous and multiparous cows. Median (range) for BCSratio were 0.91 (0.61-1.00) and 0.87 (0.53-1.00) for primiparous and multiparous cows, respectively. Predicted probabilities for PAI1 for low, medium, and high BCSratio categories were 34.1%, 38.2%, and 38.3% in primiparous and 18.2%, 24.3%, and 25.2% in multiparous cows, respectively. The logistic regression analyses identified significant associations between BCSratio and PAI1, where cows with greater BCSratio values were more likely to conceive at AI1. The analyses indicated that the odds (95% CI) of PAI1 increased by 2.47 (1.25-4.91; $P = 0.009$) and by 3.22 (1.48-7.06; $P = 0.003$) for each 0.5-unit increment in BCSratio in primiparous and multiparous cows, respectively. Overall, the magnitude of the reduction between BCS at calving and nadir BCS had a significant impact on pregnancy at first artificial insemination.

Keywords: Body condition, automated, nadir, fertility.

Introduction

The change from late gestation to early lactation is accompanied by remarkable metabolic and endocrine adjustments in the dairy cow (Gross *et al*., 2011). As this transition is characterized by a reduction in dry matter intake associated with an increased demand for nutrients to support the initiating lactation, cows experience a typical negative energy balance (NEB) during the peripartum period. Fat and labile protein mobilization from body energy reserves to match the increased energetic demands results in reductions in subcutaneous fat which can be assessed by body condition scoring (BCS) (Lean *et al*., 2013; Roche *et al*., 2009).

In practical terms, dairies watch for body condition to monitor cows' energy status and energy balance. Nonetheless, only one third of the US dairy farms implemented formal BCS into their management practices (Hady *et al*., 1994; Bewley *et al*., 2010), likely due to the time consuming and subjective nature of visual or tactile assessment (Edmondson *et al*., 1989; Leroy *et al*., 2005).

The advent of automated body condition scoring systems has allowed for the use of data originated at multiple and precise time points, with scores that are not affected by inter and intra evaluator variation (Borchers and Bewley, 2015).

Previous research has identified the impact of inadequate energy status and energy balance on cow fertility (Carvalho *et al*.; 2014; Roche *et al*., 2009; Barletta *et al*., 2017). Nonetheless, these studies were completed considering visual body condition evaluation at specific time points, which makes it difficult an evaluation of the magnitude and timing of the largest reduction in BCS during early lactation. Availability of daily scores originated from automated camera systems provides opportunity for precise assessment of the impact of BCS nadir on subsequent cow performance.

We hypothesized that the magnitude of the reduction between BCS at calving and nadir BCS would have a significant impact on< pregnancy at first artificial insemination. Therefore, the objective of this study was to evaluate the effect of the decrease in BCS from calving to nadir BCS before first artificial insemination (AI1) on pregnancy per AI1 (PAI1) in Holstein cows.

Material and methods

Study design and study population

This retrospective observational study included information collected from 6,100 lactations (primiparous = 3,683; multiparous = 2,417) starting between April 2019 and March 2021 in a commercial dairy operation located in Colorado, USA. Cows were maintained in a cross ventilated barn, milked 3X in a 90 units rotary parlor, and subject to first AI at about 80 DIM (primiparous) and 60 DIM (multiparous), following a Double OvSynch protocol. Pregnancy diagnosis was performed via transrectal ultrasonography on d 32±3 after AI and reconfirmed at d 80±3 of gestation. Cows determined non pregnant were administered prostaglandin F2α if a corpus luteum was visible and were submitted for AI based on estrus detection using the DeLaval activity meter system (DelPro Farm Manager software DeLaval International AB, Tumba, Sweden).

Data collection started at dry-off (multiparous) or at calving (primiparous) and continued until the AI resulting in pregnancy or culling. Cow demographic, reproductive, and health data were extracted from on-farm software (Dairy Comp 305; Valley Ag Software, Tulare, CA). Daily milk yield and BCS were extracted from DelPro Farm Manager software. The dataset included cow ID, date of calving, lactation number, calving-related and disease events, breeding dates, pregnancy diagnosis outcomes, daily milk yield for the first 60 DIM, and daily BCS.

Body condition scoring and BCS categorization

Scores were generated by an automatic BCS system using DeLaval BCS cameras (DeLaval International AB, Tumba, Sweden) previously validated by Mullins *et al*. (2019) that were mounted on the sorting-gate at each exit (n = 2) of the milking parlor. As the cow passed under the mounted camera, a continuous video (30 FPS, 32,000 captured reference points) was taken and a 3D image from the video was automatically created and saved by the BCS camera software (Mullins *et al*., 2019; Pinedo *et al*., 2022). In a secondary step, the saved 3D images were processed through an algorithm and analyzed to locate the key physical characteristics (pins, tail head ligaments, sacral ligaments, short ribs, and hooks) of the cow to calculate the automated BCS, viewable in DelPro Farm Manager. The proprietary algorithm used the BCS scoring scale proposed by earlier studies, modified to report BCS in 0.1-point increments (Ferguson *et al*., 1994).

All automated BCS data were recorded in and downloaded from DelPro Farm Manager and scores generated by BCS cameras at calving (BCScalv) and nadir (BCSnadir) were selected to calculate the ratio BCS at nadir/BCS at calving (BCSratio). The BCSratio is a representation of the BCS change from calving to nadir, where greater BCS loss results in smaller values for the ratio.

Calving-related events and disease events were obtained from farm records stored in on-farm software. Only health events diagnosed before or at the day of AI1 were considered in the analyses. Parity was created as a binary variable including primiparous (lactation number =1) and multiparous (lactation number ≥2) cows. Calvings were grouped by season (spring, summer, fall, or winter). Finally, a milk yield category was added as a covariable in the models using the quartile distribution of the average daily milk yield in the first 60 DIM (**M60**) obtained from DelPro Farm Manager.

Statistical analyses

All the analyses were performed separately for primiparous and multiparous cows. Descriptive time-to-event analysis for pregnancy was performed using PROC LIFETEST in SAS 9.4 (SAS institute Inc., Cary, NC).

Initial univariable models using only BCSratio as explanatory variable were followed by multivariable models that considered calving season, occurrence of disease up to AI1, and milk yield up to 60 DIM as covariables. Descriptive statistics were calculated using the PROC UNIVARIATE. Least square means for BCS and for days to AI1 by parity category were calculated and compared using ANOVA (PROC GLM).

Odds ratios (**OR**) and predicted probabilities for pregnancy at AI1 were estimated for the explanatory variables of interest using PROC GLIMMIX. For all outcome variables, significant predictors were selected at *P*-value <0.05; interaction terms and controlling variables remained in the models at *P*-value ≤0.10.

Results and discussion

The analysis included 6,100 lactations (primiparous = 3,683; multiparous = 2,417). Overall, distribution of calvings across seasons were spring 15.9%, summer 36.2 %, fall 27.8%, and winter 20.1%. Average milk yield for the first 60 DIM was 30.4 (0.07) kg and 45.3 (0.06) kg for primiparous and multiparous cows, respectively.

Median time from calving to nadir were 51 d and 59 d for primiparous and multiparous cows, respectively, while average BCS change from calving to nadir in primiparous and multiparous cows were -0.22 and -0.35 (Table 1). Median (range) for BCSratio were 0.91 (0.61-1.00) and 0.87 (0.53-1.00) for primiparous and multiparous cows, respectively. These values agree with a recent report that indicated that days to nadir were 38 d and 54 d in primiparous and multiparous cows housed in a large commercial dairy farm in Indiana (Truman *et al*., 2022). In the same study, primiparous cows lost 0.14 BCS points, while multiparous cows lost 0.30 points from calving to nadir BCS.

Mean (SD) DIM to AI1 for primiparous and multiparous cows were 83.3 (11.5) d and 63.4 (11.5) d (*P* <0.001) and pregnancy per AI1 was 33.9% (primiparous = 42.2%; multiparous = 28.1%; *P* <0.001). Predicted probabilities for PAI1 for low, medium, and high BCSratio categories were 34.1%, 38.2%, and 38.3% in primiparous and 18.2%, 24.3%, and 25.2% in multiparous cows, respectively.

Data for this study originated from an automated BCS system, which allowed for the identification of the BCS nadir through daily measurements. Thus, the system provides the exact magnitude of the BCS reduction during early lactation. The nadir calving ratio is a representation of this BCS change, where smaller values indicate more severe BCS loss.

The logistic regression analyses identified significant associations between BCSratio and PAI1. As expected, cows with greater BCSratio values were more likely to conceive at AI1. The analyses indicated that the odds (95% CI) of PAI1 increased by 2.47 (1.25-4.91; *P* = 0.009) and by 3.22 (1.48-7.06; *P* = 0.003) for each 0.5-unit increment in BCSratio in primiparous and multiparous cows, respectively.

Previous studies have reported the associations among BCS variables and multiple reproductive variables (Roche *et al*., 2007; Carvalho *et al*., 2014; Chebel *et al*., 2018). As reviewed by Roche *et al*. (2009), most of the reports studying the physiological effects of energy status and energy balance on fertility suggest a positive association between an earlier achievement of pregnancy and increased BCS and reduced BCS loss during early lactation. The results from the current study align with those reported recently by our group in a similar analysis where cows with large loss in BCS between calving to 21 DIM, 56 DIM, and AI1 had lesser odds of P/AI1 compared with other categories of ÄBCS within the same time period (Pinedo *et al*., 2022). The strength of our study is the availability of daily BCS for a detailed description of the BCS drop from calving to nadir. To our knowledge, this is the first study testing the effects of BCS nadir on pregnancy outcomes in a large population of dairy cows. In primiparous cows, insemination started on average 30 days after nadir, at a time where we expect their energy level to be more in balance. In contrast, inseminations for multiparous cows started on average at the same time as they reached nadir. This could be an important factor to explain the lower achieved P/AI1 in multiparous cows. Further studies should investigate cow and parity specific best breeding windows based on BCS profiles.

The reasons for the reduced fertility in cows losing BCS are not fully understood. Nonetheless, the follicular/oocyte quality could be affected by suboptimal energy status during the early postpartum, impairing subsequent fertility in lactating dairy cows (Britt,

Table 1. Descriptive statistics for body condition scores, days in milk at nadir, and average milk yield in the first 60 DIM. Unless stated, least square means (SE) are presented.

Parameter	Primiparous	Multiparous
BCS at calving	3.38 (0.009)	3.34 (0.009)
BCS at nadir	3.15 (0.009)	2.99 (0.009)
BCS change	-0.22	-0.35
DIM at nadir (d)	51 (0.18)	59 (0.21)
BCSratio (median [range])[1]	0.91 (0.61-1.00)	0.87 (0.53-1.00)
Milk yield (average 60 DIM; kg)	30.4 (0.07)	45.3 (0.06)

[1] BCSratio = Calculated as BCS at nadir/BCS at calving.

1992). In addition, changes in hormone levels that regulate gene expression and the secretion of proteins by the endometrium, could affect implantation or pregnancy recognition (Beam and Butler, 1999).

Conclusion

Overall, the magnitude of the reduction between BCS at calving and nadir BCS, analysed here as BCSratio, had a significant impact on pregnancy at first artificial insemination. Larger reductions in BCS were associated with lower likelihood of pregnancy, with a more pronounced effect in multiparous than primiparous cows.

Automatic BCS is a useful tool to monitor and manage energy balance. Individual cow or group BCS profiles should be considered to determine the onset of breeding windows. The implications for fertility management deserve further exploration.

Acknowledgement

This work is supported by Food and Agriculture Cyberinformatics and Tools grant no. 2019-67021-28823 from the USDA National Institute of Food and Agriculture. We thank Wolf Creek Dairy for allowing access to their cow records and DeLaval, International AB, Tumba, Sweden for providing the DeLaval body condition scoring cameras, as well as their assistance in the data acquisition from DelPro Farm Manager software. The authors declare that they have no competing interests. Joaquin Azocar is employed by DeLaval Inc, USA and Ilka Klaas is employed by DeLaval International AB, Sweden.

References

Barletta, R.V., M. Maturana Filho, P. D. Carvalho, T. A. Del Valle, A. S. Netto, F. P. Rennó, R. D. Mingoti, J. R. Gandra, G. B. Mourão, P. M. Fricke, R. Sartori, E. H. Madureira, and M. C. Wiltbank. 2017. Association of changes among body condition score during the transition period with NEFA and BHBA concentrations, milk production, fertility, and health of Holstein cows. Theriogenology 104:30-36.

Beam, S. W. and W. R. Butler. 1999. Effects of energy balance on follicular development and first ovulation in postpartum dairy cows. J. Reprod. Fertil. 54:411-424.

Bewley, J. M., M. D.Boehlje, A. W. Gray, H. Hogeveen, S. D. Eicher, M. M. Schutz. 2010. Assessing the potential value for an automated dairy cattle body condition scoring system through stochastic simulation. Agric. Financ. Rev. 70:126-150.

Borchers, M. R. and J. M. Bewley. 2015. An assessment of producer precision dairy farming technology use, prepurchase considerations, and usefulness. J. Dairy Sci. 98:4198-4205.

Britt, J. 1992. Impacts of early postpartum metabolism on follicular development and fertility. Proc. Annu. Conv. Am. Assoc. Bovine Pract. Am. Assoc. Bovine Pract; Auburn, AL. p. 29-43.

Carvalho P. D., A. H. Souza, M. C. Amundson, K. S. Hackbart, M. J. Fuenzalida, M. M. Herlihy, H. Ayres, A. R. Dresch, L. M. Vieira, J. N. Guenther, R. R. Grummer, P. M. Fricke, R. D. Shaver, and M. C. Wiltbank. 2014.

Relationships between fertility and postpartum changes in body condition and body weight in lactating dairy cows. J. Dairy Sci. 97:3666-3683.

Chebel, R. C., L. G. D. Mendonca, and P. S. Baruselli. 2018. Association between body condition score change during the dry period and postpartum health and performance. J. Dairy Sci. 101:4595-4614.

Edmondson, A. J., I. J. Lean, L. D. Weaver, T. Farver, and G. Webster. 1989. A body condition scoring chart for Holstein cows. J. Dairy Sci. 72:68-78.

Ferguson, J. D., D. T. Galligan, and N. Thomsen. 1994. Principal descriptors of body condition score in Holstein cows. J. Dairy Sci. 77:2695-2703.

Gross, J., H. A. van Dorland, R. M. Bruckmaier, and F. J. Schwarz. 2011. Performance and metabolic profile of dairy cows during a lactational and deliberately induced negative energy balance with subsequent realimentation. J. Dairy Sci. 94:1820-1830.

Hady, P. J., J. J. Domecq, and J. B. Kaneene. 1994. Frequency and precision of body condition scoring in dairy cattle. J. Dairy Sci. 77:1543-1547.

Lean, I. J., R. Van Saun, and P. J. DeGaris. 2013. Energy and protein nutrition management of transition dairy cows. Vet Clin N Am-Food A 29:337-366.

Leroy, T., J. M. Aerts, J. Eeman, E. Maltz, G. Stojanovski, and D. Berckmans. 2005. Automatic determination of body condition score of cows based on 2D images. Pages 251–255 in Precision Livestock Farming. S. Cox, ed. Wageningen Press, Wageningen, the Netherlands.

Mullins, I. L., C. M. Truman, M. R. Campler, J. M. Bewley, and J. H. C. Costa. 2019. Validation of a commercial automated body condition scoring system on a commercial dairy farm. Animals. 9(6);287.

Pinedo, P., D. Manríquez, J. Azocar, B. R. Klug, and A. De Vries. 2022. Dynamics of automatically generated body condition scores during early lactation and pregnancy at first artificial insemination of Holstein cows. J. Dairy Sci. 105:4547–4564.

Roche, J. R., K. A. Macdonald, C. R. Burke, J. M. Lee, and D. P. Berry. 2007. Associations among body condition score, body weight, and reproductive performance in seasonal-calving dairy cattle. J. Dairy Sci. 90:376–391.

Roche, J. R., N. C. Friggens, J. K. Kay, M. W. Fisher, K. J. Stafford, and D. P. Berry. 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. J. Dairy Sci. 92:5769–5801.

Truman, C. M., M. R. Campler, and J. Costa. 2022. Body condition score change throughout lactation utilizing an automated BCS system: A descriptive study. Animals. 12(5), 601.

Early mastitis detection: Can Lactoferrin evaluation by EU funded project MOLOKO biosensor help?

*P. Bulgarelli[1], A. Comparelli[1], F. Marabelli[2], E. Manobianco[2], L. Lopez Sanchez[2],
S. Toffanin[3] and J. Peters[4]*

*[1]Parmalat S.p.A., Via delle Nazioni Unite, 43044 Collecchio, Italy
Corresponding Author: paolo.bulgarelli@it.lactalis.com
[2]Plasmore Srl, Viale Vittorio Emanuele II 4, 27100 Pavia, Italy
[3] CNR-ISMN, Institute for the Study of Nanostructured Materials, Italian National Research Council, Via Gobetti 101, 40129 Bologna, Italy
[4] WFSR, Wageningen University and Research, Akkermaalsbos 2,
6708 WB Wageningen, The Netherlands*

Abstract

Mastitis is the most frequent udder disease in dairy cows. It causes huge economic losses, due to less milk production, less milk quality, increasing of drugs usage and, in most severe cases, early culling.

A healthy herd is a fundamental goal not only for animal welfare but also for the environment. Mastitis reduces herd efficiency by the increasing of environmental impact of milk production. In fact, it was calculated that a sick cow produces 6.3 % CO_2 eq./kg milk more than a healthy one.

To reach efficiency goals, Precision Livestock Farming can give an important contribution. In compliance with that, MOLOKO (Multiplex phOtonic sensor for pLasmonic-based Online detection of contaminants in milK) European funded project aims development of a biosensor to improve milk safety and animal health.

The MOLOKO biosensor working mechanism is based on the optical measurement of an inherent physical property change of the environment in the proximity of the sensing surface.

Lactoferrin was chosen as a mastitis biomarker due to its antimicrobial activities. It is an iron-binding glycoprotein synthesized by neutrophilic polymorphonuclear leukocytes and granular epithelial cells in milk and other exocrine secretions.

A direct immunoassay for the detection of Lactoferrin was developed in buffer and in diluted raw milk using the MOLOKO sensor. The test showed a correct and cross-correlated recognition of Lactoferrin on suitably functionalized channels; promising results were obtained for both buffer (Limit Of Detection ~ 9µg/ml) and diluted raw milk (Limit Of Detection ~40µg/ml) providing information regarding levels of Lactoferrin in only 14 minutes.

A preliminary validation of the sensor for the detection of Lactoferrin was carried out with milk samples collected from two different dairy herds in the north of Italy, respectively 600 milking cows, with a milking parlour and 100 milking cows with automatic milking system.

Samples analysed in the MOLOKO sensor and by certified laboratory were compared for the estimation of lactoferrin concentration. Additionally, known parameters of udder infections, such as somatic cells count, differential somatic cells count, and finally bacteriologic culture were also determined to better identify true positive.

The preliminary data show that integration of the MOLOKO biosensor on the milking systems could be useful to detect daily lactoferrin fluctuations. Combined with other sensors, it could provide farmers information about inflammatory events.

Keywords: MOLOKO, biosensor, precision livestock farming, mastitis, lactoferrin.

Introduction

Mastitis is the most frequent udder disease in dairy cows. It causes huge economic losses, due to less milk production, less milk quality, increasing of drugs usage and, in most severe cases, early culling (Halasa *et al*., 2011).

Worldwide, was estimated that mastitis costs between 61€ to 97€ per cow with clinical case and between 17€ to 198 € per cow with subclinical one (Hogeveen, 2011).

A healthy herd is a fundamental goal not only for animal welfare but also for the environment. Mastitis reduces herd efficiency by the increasing of environmental impact of milk production. In fact, it was calculated that a sick cow produces 6.3 % CO_2 eq./kg milk more than a healthy one (Mostert *et al*, 2019).

To reach efficiency goals, Precision Livestock Farming can give an important contribution; indeed, Precision Livestock Farming allows to monitor every animal in real-time using sensors that check their health, welfare, production and reproduction status (Berckmans, 2017).

In compliance with that, MOLOKO (Multiplex phOtonic sensor for pLasmonic-based Online detection of contaminants in milK) European funded project aims development of a biosensor to improve milk safety and animal health.

One of the aspects that MOLOKO biosensor evaluates is mastitis risk; to allow early mastitis detection and improve health, productivity, and welfare of dairy cattle, understanding immune factors and mechanisms involved in the mammary gland defence against infection represents a fundamental task (Chaneton *et al*., 2013).

In the bovine mammary gland, the viable leukocytes offer some degree of cellular protection against microbial invasion through their ability to recognise microorganisms and induce a rapid inflammatory response in an attempt to resolve the intra mammary infection immediately (Alnakip *et al*.,2014).

Lactoferrin (LF) is an iron-binding glycoprotein, found in milk and other exocrine secretions which has antimicrobial/antiviral activities, immunomodulatory activity and antioxidant activity (Wakabayashi *et al*., 2006); LF is synthesized by neutrophilic polymorphonuclear leukocytes that are the first recruited immune cell to sites of infection (Alnakip *et al*., 2014), (Figure 1).

Furthermore, it has been demonstrated that LF concentration changes according to stage of lactation, parity and presence of pathogen, in fact it is significantly higher in cows with subclinical mastitis than healthy ones (Hagiwara *et al*., 2003).

For these reasons Lactoferrin was chosen as a MOLOKO mastitis biomarker.

The aim of this study was to assess MOLOKO biosensor for LF detection in raw cow milk and compare it, as a mastitis biomarker, with a gold standard LF detection.

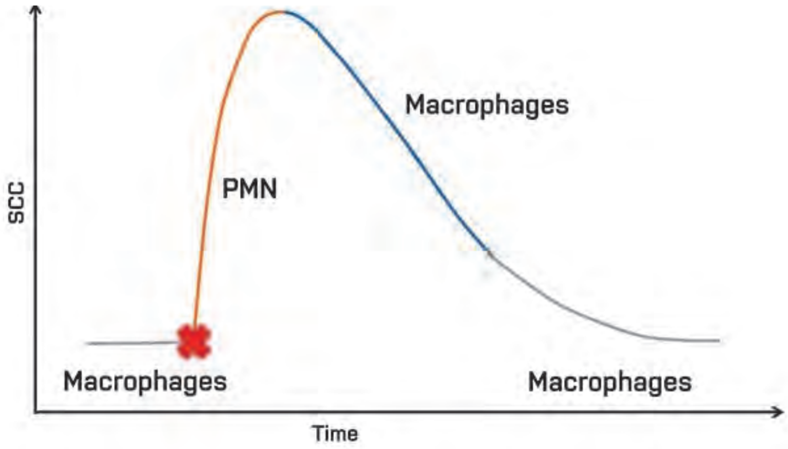


Figure 1. Udder immunologic response after infection (Alnakip et al., 2014).

Material and methods

MOLOKO is a European Horizon 2020 funded project, comprising 12 partners from across Europe, with the aim to optimise production and safety throughout the milk supply chain through the development of a fast and cheap biosensor; its working mechanism is based on the optical measurement of an inherent physical property change of the environment in the proximity of the sensing surface (Figure 2).

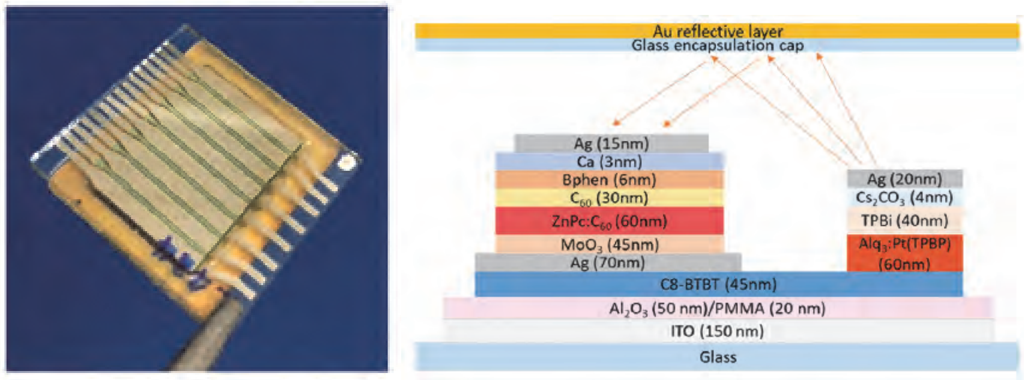


Figure 2. Bottom side of the photonic module. Prosa et al. (2021), Organic Light-Emitting transistor in a Smart-Integrated system for Plasmonic-Based sensing.

MOLOKO partners have decided to research both potential contaminants in milk (e.g. antibiotics, mycotoxins, bacterial toxins), and at the same time create a diagnostic tool useful for the reduction of antibiotics prescriptions, responding to the concerns of WHO and European Commission regarding overuse of antibiotics and subsequent drug resistance.

A direct immunoassay for the detection of Lactoferrin was developed in buffer and in diluted raw milk using the MOLOKO sensor.

The test showed a correct and cross-correlated recognition of Lactoferrin on suitably functionalized channels; promising results were obtained for both buffer (Limit Of Detection ~ 9µg/ml) and diluted raw milk (Limit Of Detection ~40µg/ml) providing information regarding levels of Lactoferrin in only 14 minutes.

Individual milk samples were collected from 2 dairy herds in the North of Italy, from January 2022 to March 2022; these herds were very different: they had respectively 600 milking cows, with a milking parlour and 100 milking cows with automatic milking system.

Cows were selected randomly according to all factors can influence LF concentration, such as stage of lactation, daily milk production, SCC, and parity; all this information, at the beginning of sampling were recorded.

Milk samples were collected in 2 refrigerated aliquots, first one was used for MOLOKO biosensor analysis and the second one was delivered to certified laboratory.

Before analysis with MOLOKO biosensor, samples preparation was needed; milk samples were diluted 1:100 and were filtered twice, first one with 0.45 µm filter and second one with 0.2 µm filter.

ELISA test was performed as a gold standard by certified laboratory in the north of Italy; kit used was "5091LFER- EuroProxima Lactoferrin: lactoferrin in various matrix".

After analysis the colour intensity of the samples were read automatically and the concentration of lactoferrin were calculated.

Additionally, known parameters of udder infections, such as somatic cells count, differential somatic cells count, and finally bacteriologic culture were also determined by certified laboratory and other partner of MOLOKO Project.

Results and discussion

Preliminary analyses were made both in buffer and milk; after this analysis, we could observe a clear signal of 0.014% in the specific channel for the detection of lactoferrin in buffer (Figure 3) which indicated that the OPM was responding well to the detection of lactoferrin.

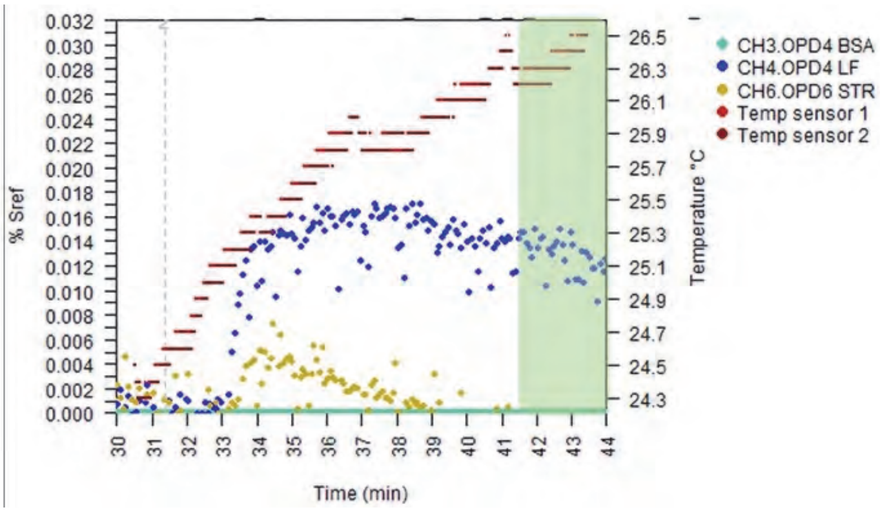


Figure 3. Sensorgram of the interaction of Lactoferrin at 100 µg/ml in buffer.

In the sensorgram for the milk samples diluted 100 times (Figure 4) and after dissociation (green region), we did not observe an increasement in the signal respect to the baseline. That indicates that the concentration of Lactoferrin in those samples is below the limit of detection of this OPM (approximately 0.001% which would correspond to 10 μg/ml of Lactoferrin in buffer and 1000 μg/ml in milk sample).

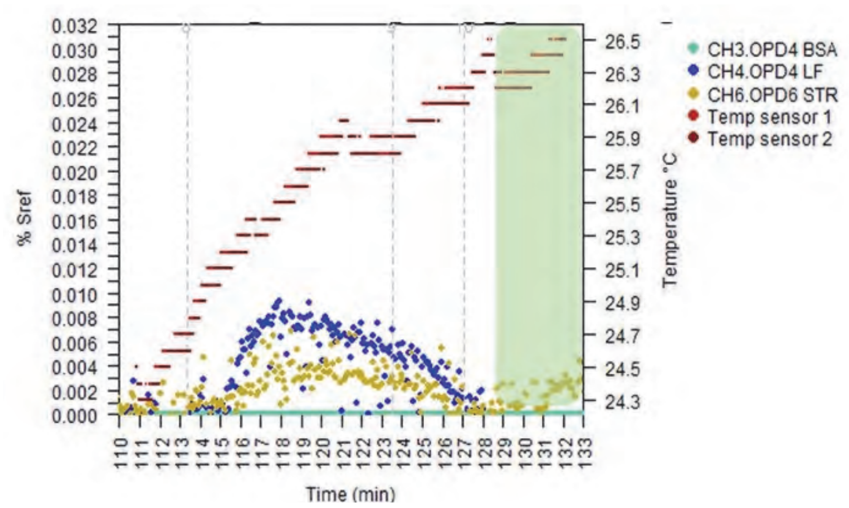


Figure 4. Sensorgram of the interaction of milk sample diluted 100 times.

However, we can say that qualitatively the concentration of Lactoferrin in milk samples was lower than 1000 μg/ml, according to ELISA test the level was 361 μg/ml.

Kutila T. (2004) have reviewed sundry studies and have reported main results about milk LF concentration in healthy and mastitic cows:

- Lactating cow: 20-35 μg /ml.
- Cow with subclinical mastitis: 20-120 μg /ml.
- Cow with clinical mastitis: 20-230 μg /ml.

With data collected from the sampling, further assessment will help in confirming the given threshold especially due to LF influence by parities and different stage of lactation (Hagiwara *et al*., 2003), also bacterial presence will be assessed.

Conclusion

The preliminary data showed that LF is correctly recognized by MOLOKO biosensor; for this reason, its integration on the milking systems could be useful to detect daily lactoferrin fluctuations. Combined with other sensors, it could provide farmers information about inflammatory events, furthermore, another fundamental goal it could be define a more accurate LF thresholds to distinguish clinical mastitis from subclinical ones and effectiveness at drying off.

However, more study is needed to define a sample preparation that is consistent with lactoferrin level in raw milk.

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 780839 (MOLOKO Project) www.moloko-project.eu

References

Alnakip M. E., Quintela-Baluja M., Bohme K., Fernandez-No I., Caamano-Antelo S., Calo-Mata P., Barros-Velaquez J. (2014), The immunology of mammary gland of dairy ruminants between healty and inflammatory conditions, Journal of Veterinary Medicine, Vol. 2014.

Berckmans (2017). General introduction to precision livestock farming. Animal frontiers, Jan. 2017, Vol. 1, No. 1.

Chaneton L., Bontà M., Pol M., Tirante L., Bussmann E. (2013), Milk lactoferrin in heifers: influence of health status and stage of lactation, Journal of Dairy Science Vol. 96 No. 8, 2013, 4977-4982.

Hagiwara S., Kawai K., Anri A., Nagahata H. (2003), Lactoferrin Concentrations in Milk from Normal and Suclinical Mastitic Cows, Journal of Veterinary Medical Science - March 2003.

Halasa T., Huijps K., Osteras O., and Hogeveen H. (2011). Economic effects of bovine mastitis and mastitis management: A review. Veterinary Quarterly.

Hogeveen H. (2011), Economic aspects of mastitis: New developments, New Zealand Veterinary Journal, Volume 59, 2011.

Kutila Taina (2004), Role of Lactoferrin in treatment of Bovine Mastitis.

Mostert P. F., Bokkers E. A. M., De Boer I. J. M., Van Middelaar C. E. (2019), Estimating the impact of clinical mastitis in dairy cows on greenhouse gas emissions using a dynamic stochastic simulation model: a case study, The International Journal of Animal Biosciences, Vol. 13, Issue 12 (2019), 2913-2921.

Prosa P., Benvenuti E., Kallweit D., Pellacani P., Toerker M., Bolognesi M., Lopez-Sanchez L., Ragona V., Marabelli F., Toffanin S. (2021), Organic Light-Emitting Transistor in a Smart-Integrated System for plasmonic-Based Sensing, Advanced Functional Materials, 2021, 2104927.

Wakabayashi H., Yamauchi K., Takase M. (2006), Lactoferrin research, technology and applications, International Dairy Journal 16 (2006) 1241-1251.

Challenges of integration and validation of farm and sensor data for dairy herd management

K. Schodl[1,2], B. Fuerst-Waltl[1], H. Schwarzenbacher[2], F. Steininger[2], M. Suntinger[2], F. Papst[3,4], O. Saukh[3,4], L. Lemmens[5], D4Dairy-Consortium[2] and C. Egger-Danner[2]

[1]University of Natural Resources and Life Sciences, Vienna, Gregor-Mendel-Strasse 33, 1180 Vienna, Austria
[2]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Strasse 89/B1/18, 1200 Vienna, Austria
[3]Graz University of Technology, Rechbauerstrasse 12, 8010 Graz, Austria
[4]Complexity Science Hub, Josefstaedter Strasse 39, 1080 Vienna, Austria
[5]University of Veterinary Medicine Vienna, Veterinaerplatz 1, 1210 Vienna, Austria

Abstract

Precision livestock management has become an integral part of agriculture and dairy farms are increasingly using precision technologies such as sensor systems for herd health monitoring. Various companies exist, offering different sensors, e.g. accelerometers in collars or boli measuring cow activity and rumination, for different purposes such as fertility, health monitoring or feeding management. The D4Dairy project aimed, amongst others, to investigate the potential of sensor data and other farm and cow-specific data for disease prediction and genetic improvement of metabolic, udder and claw health. Results should lay a foundation for herd management tools and genetic health indices, which are expected to work across farms and sensor systems. Prior to any application, validation of sensor measurements is necessary. Some companies validated their sensor technologies in scientific studies by comparing sensor measurements (e.g., rumination time, duration of lying bouts) to a gold standard such as direct or video behavioral observations. Another aspect comprises validation of changes or patterns in sensor variables for a desired outcome such as heat or health alarms. Furthermore, for the implementation of monitoring of any kind, reliability of measurements is another crucial aspect. Erroneous measurements due to hardware or software malfunctioning have to be identified correctly and outliers have to be distinguished from true deviations. The latter is even more difficult for sensor measurements without possibilities for a plausibility check. Activity indices or other dimensionless sensor outputs lack established reference values and may differ even between animals equipped with the same sensor type whereas plausibility of rumen temperature or milk yield can also be assessed based on empirical knowledge. These issues had and partly still have to be overcome in the D4Dairy project and all projects with similar aims. In our contribution we want to present our approaches to sensor data validation, the problems we encountered and how we dealt with them including general recommendations for future studies in this area.

Introduction

Technologies have advanced on dairy farms during the last decades, which has released a potential for precision livestock farming. One of these advances is sensor technology for dairy cows for fertility and herd health management. Various technologies exist measuring for example activity, rumination, or reticular temperature to identify cows in heat for insemination, cows, which are about to give birth or send health alarms. These notifications are based on continuous recording and evaluation of these parameters based on changes in activity, rumination, or temperature patterns in the individual animal. Sensors recording these data are administered to the animal using for example collars, nosebands, foot bands, or rumen boluses. Crucial, of course, is that the sensor device is linked to the individual animal by a unique ID.

Many manufacturers already offer products using sensor technology for detecting cows in heat or at the onset of calving or cows, which may need medical treatment. However, sensor data is often very noisy and before these technologies yield reliable results it is important to clean and validate the (raw) data. There are different ways to approach sensor data validation, depending very much on the purpose and its area of use. The most obvious and sound approach is the comparison to a gold standard such as behavioral observations to assess the sensor's precision (Grinter *et al*., 2019). Due to the high time expenditure of behavioral observations other approaches relied on the comparison against other, already validated devices (Elischer *et al*, 2013) or the agreement between two devices on the same animal (Kok *et al*, 2015). Stygar *et al*. (2021) reviewed if and how various sensor systems offered on the market have been validated.

Aside from the validation of sensor technology for the target customers, these devices also offer a great opportunity to be used for research or other fields of application such as phenotyping for routine genetic evaluation or the development of new decision support tools for farmers. However, it is crucial for any user of the sensor data to know if and how data from these sensors have been validated and how much the data has been pre-processed prior to provision. Commonly, sensor systems available on the market are validated for one or a few specific purposes. Thus, data processing and software algorithms may emphasize some behaviors more than others to generate the most reliable alarms, for example emphasizing mounting activity for heat detection (Elischer *et al*., 2013). Depending on the use of these data for research or other purposes it may be necessary to additionally validate the sensor data or algorithms.

One of the aims of the D4Dairy project (https://d4dairy.com/) was to investigate the potential of sensor derived data and their integration with other farm data, such as veterinary records and diagnoses or data from automatic milking systems (AMS) for disease prediction and genetic improvement of metabolic, udder and claw health. For this purpose, farms already using sensor technology were motivated to participate in data collection. The sensor systems used on these farms were the rumen bolus by smaXtec (smaXtec animal care GmbH; 25 farms), the Lely T4C system (Lely International N.V., 35 farms), the SenseHub^{TM} Dairy system (Allflex Livestock Intelligence, MSD Animal Health, 10 farms), the DelPro^{TM} Farm Manager system by DeLaval (14 farms), and CowScout and Rescounter by GEA (9 farms). Additional information was derived from milking systems, veterinary records and diagnoses, national performance recordings, breeding information (e.g., genomic data, estimated breeding values), farm records and information on the operational structures on the farms, management information, climate sensor systems and weather stations, claw trimmings, rapid blood and milk tests for ketosis, and BCS and lameness recordings. Results should lay a foundation for herd management tools and genetic health indices, which are expected to work across farms and sensor systems. However, prior to any further analysis, sensor data had to be inspected and evaluated carefully so results would not be biased by erroneous data due to sensor malfunctioning or measurement errors. In this paper we describe how we approached data validation for the sensor systems by smaXtec, Lely, and Allflex as well as data from AMS and draw conclusions for automatization of data cleaning pipelines for future applications.

Animals, material, and methods

Sensor data were collected between January 2019 and August 2021 on farms with smaXtec sensors, between January 2020 and March 2021 on farms with the system by Lely and between January 2020 and May 2021 on farms with the SenseHub^{TM} sensor system. The smaXtec bolus measured activity and temperature in the cow's reticulum and information was read from the sensors in intervals of ten minutes. The activity data was provided as a dimensionless number prior to any handling by the manufacturer whereas temperature was available in a raw format as well as after correction for temperature drops caused by drinking, which was further used for analysis. The other two sensor systems measured activity and rumination and were provided in a more aggregated format hourly (Allflex, Lely) or every two hours (Lely). Whereas Lely only provided one (dimensionless) number for activity, the data from the Allflex system was available as minutes of activity per hour for different activity levels (resting, medium activity, or high activity) and as an index called 'activity trend'. Rumination was either provided as minutes during the last 24 hours (Lely) or minutes per hour (Allflex).

Furthermore, data of daily milkings was provided by the Austrian central cattle database (Rinderdatenverbund – RDV) for single milkings as well as aggregated over 24 hours over the respective time period.

Approaches to data cleaning and outlier detection

Outliers deviate from the majority of a sample; however, they are difficult to identify in high dimensional data (Paulheim and Meusel, 2015) and it is challenging to clearly distinguish outliers from extreme deviations, which may even be of particular interest for a research question. For the smaXtec data set different approaches for outlier detection were taken up. The first one applied Isolation Forest, an unsupervised machine learning approach for outlier detection building on the co-dependency of data quality and model robustness (Papst *et al*., 2021): outliers can be identified based on an Isolation Forest Score derived from Isolation Trees, which are created during model training. This score indicates how likely it is for a data point to be an outlier (Papst *et al*., 2021). The second approach used classical plausibility checks based on domain expert knowledge and basic statistical approaches to data cleaning. This comprised the exclusion of duplicates and missing values as well as potential measurement errors with measurements occurring multiple times within the regular time window of data retrieval or where time to previous and following measurement exceeded the regular frequency. Whereas further plausibility checks of maximum and minimum values in the data were somewhat possible for temperature values using a priori knowledge on body temperature of cattle, this was not the case for the arbitrary activity value. Temperature values in the data set ranged from -42.8°C to 42.8°C. As negative temperature values are a physiological impossibility in warm-blooded animals this clearly indicated faulty data. However, defining a clear cut-off value for plausible temperature values was not straightforward and thus it was decided to use maximum deviation of three standard deviations from the overall temperature mean as a threshold. This yielded a plausible temperature spectrum between 36.6°C and 42.8°C for the whole data set (Figure 1).

As already mentioned before, there was no reference to assess plausibility of activity values and thus these were validated based on the associated temperature values. Additionally, if activity values were zero during at least twelve in 24 hours, these days were excluded because the sensor was probably not yet administered to the animals. All in all, approximately 5% of the smaXtec sensor data were discarded based on these decision criteria.

Sensor data by Lely and Allflex have been pretreated more intensively prior to provision and thus it was not known if and based on which criteria data have already been removed or altered beforehand. Although the SenseHub^{TM} system also provided an arbitrary activity index, all other parameters on activity and rumination were available

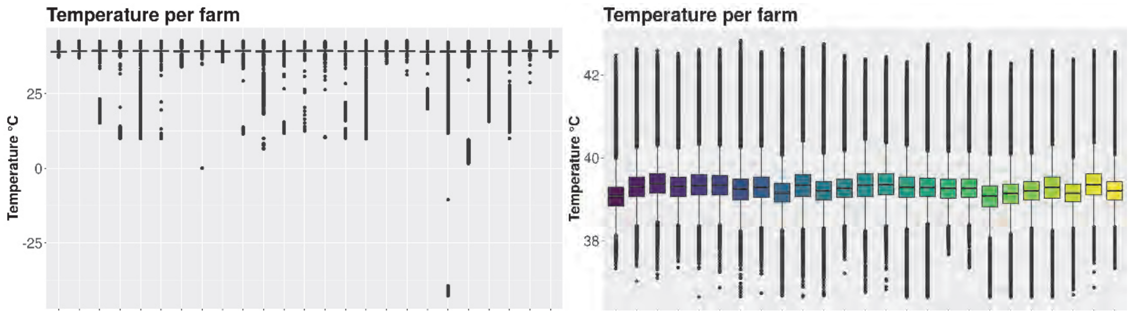


Figure 1. Reticular temperature per farm before (left side) and after (right side) data cleaning.

as minutes of this behavior per hour. Thus, if adding up all sensor variables yielded 60 minutes of behavior in one hour, data were considered correct. Furthermore, data were discarded if daily sums for rumination, eating and activity were zero for at least 24 hours suggesting that sensors were not administered, lost, or cows were removed from the herd and the sensor was not deactivated. The first day after sensor installation was also removed from the data set.

Lely sensor data either provided activity and rumination data in two-hour intervals and eating and rumination data hourly, respectively. Whereas rumination and eating were indicated in minutes of the last 24 hours, activity was again provided as dimensionless number for each two-hour interval. According to informations from the manufacturer sensors needed seven days to build a history for the animal and to reliably generate heat alarms. Thus, the first seven days of data after sensor installation were discarded. Due to the aggregation of rumination and eating time it was not easy to identify periods of potential sensor malfunctioning shorter than 24 hours. Thus, it was decided to remove data 24 hours before and after records indicating less than 10 minutes of ruminating or eating, respectively. This resulted in ranges of 10 to 893 minutes and 10 to 808 minutes for rumination and eating time, respectively. Furthermore, days with less than 11 and 22 measurements per day, respectively, were also removed from the data to avoid bias when performing further aggregation steps. Finally, four percent of the data had to be discarded based on these criteria.

Last but not least, data retrieved from AMS were validated and potential outliers were flagged including the reason. Criteria for outlier flags were the first milking of the lactation, milking intervals lower than 60 minutes and exceeding 24 hours, single milkings below one kilogram, and an hourly milk yield 50% above the ±10-day average (based on Hogeveen *et al*., 2001). Approximately 2% of the data did not fulfil these criteria. Furthermore, AMS data was matched with calving dates retrieved from the RDV to cross-validate lactation start and days in milk.

Discussion

Using commercially available sensor technology for research purposes enabled the inclusion of many farms and data due to their great availability compared to customized sensor technology in research settings. However, some limitations have to be considered such as the lack of knowledge about data processing by the sensor manufacturers prior to data provision due to trade secrets (Papst *et al*., 2019). Within the D4Dairy project this was not an issue because sensor data were not used to measure and interpret cow behavior itself. Rather, it was aimed at investigating how these sensor data may be helpful for early detection of diseases or their potential as auxiliary traits in breeding without interpretation of any physiological or behavioral relationships. Still, data had to be inspected and validated regarding measurement errors or other sources of faulty data (e.g., cows losing a sensor or devices being removed without being deactivated in the system).

All sensor types in this study were measuring activity, which was represented as a number of an undefined unit in all cases. Apart from noticeably long periods of zero activity, which indicated that the sensor must have been detached from the animal, there was no reference available for plausibility assessment. However, by using a priori knowledge about ruminal temperature as well as rumination and feeding time in cows, decisions for plausibility of a whole sensor record (activity and temperature and rumination or feeding time, respectively) were made based on the plausibility of these parameters.

Only few studies validating sensor devices for recording of rumination or using sensor measured rumination for heat, calving or disease detection provide precise information on data cleaning and outlier detection for sensor data. Reith *et al*. (2014) investigated heat detection using similar sensor systems for rumination recording like in this study and excluded values below 180 minutes and above 660 minutes per day. In her review, Beauchemin (2018) concluded that rumination and eating times range between 2.5 – 10.5 and 2.4 – 8.5 hours per day, respectively. In the present study these time ranges comprise more extreme values. However, whereas Reith *et al*. (2014) were interested in heat detection, the studies in D4Dairy aim at disease detection and lower rumination or feeding times may indicate a physiological response of unhealthy cows. The high values of rumination time observed in this sample may be due to the sensor type. According to Beauchemin (2018), substantially higher rumination times were recorded by acoustic sensors, which were also used in the present study, compared to other technologies. Furthermore, validation studies for acoustic rumination recording systems yielded good overall accordance to behavioral observations, but varied considerably between individual animals due to e.g., muscle or skin thickness or interference by background sounds (Beauchemin, 2018). Thus, this sensor system may have limited value for studies aiming at investigating rumination behavior itself, whereas disease detection based on the assessment of relative changes of patterns in individual animals may benefit a lot from this technology.

Data validation based on Isolation Forest is a promising method for identifying potential outliers. By assigning a score according to the likelihood of a data point to be an outlier, the final decision of excluding or including data may still be taken by the user, if intended. The main intention of this concept builds on the co-dependency of input data quality and model robustness to assess performance of predictive models given distribution shifts in incoming data (Papst *et al*., 2021). Whereas this data-driven approach focuses more on an application in the field, the second validation approach based on domain expert knowledge may be more suitable for ‘upstream’ research work on model formulation and feature definition.

Conclusion

Data validation and quality assurance is a crucial aspect when analyzing high dimensional data such as data from dairy cattle sensor systems. Even more so, steps of data cleaning should be comprehensive and made transparent if used for research purposes, which is not often the case in scientific literature using sensor data for device validation or further detection purposes. Limitations of data, which has been altered prior to data provision and which has been validated for specific purposes, should be taken into account in particular when using commercially available devices where processing of raw data is not known due to trade secrets. However, when being aware of these limitations these data offer a huge potential for the use in research for disease detection as well as the development of applications such as decision support tools or phenotyping strategies for auxiliary traits. Finally, integrating sensor data with other farm- and cow-specific data enables cross-validation between data sets and thus may help to additionally refine data for implausible values.

Acknowledgement

This work was conducted within the COMET-Project D4Dairy (Digitalization, Data integration, Detection and Decision support in Dairying; Project number 872039), which is supported by BMK, BMDW and the provinces of Lower Austria and Vienna in the framework of COMET – Competence Centers for Excellent Technologies. The Comet program is handled by the FFG grant number 872039.

References

Beauchemin, K.A. 2018. Invited review: Current perspectives on eating and rumination activity in dairy cows. J. Dairy Sci. 101:4762–4784.

Elischer, M.F., M.E. Arceo, E.L. Larcher and J.M. Siegford, 2013. Validating the accuracy of activity and rumination monitor data from dairy cows housed in a pasture-based automatic milkling system. J. Dairy Sci. 96:6412-6422.

Grinter, L.N., M.R. Campler and J.H.C. Costa, 2019. Technical note: Validation of a behavior-monitoring collar's precision and accuracy to measure rumination, feeding, and resting time of lactating dairy cows. J. Dairy Sci. 102:3487-3494.

Hogeveen, H., W. Ouweltjes, C.J.A.M. de Koning and K. Stelwagen, 2001. Milking interval, milk production and milk flow-rate in an automatic milking system. Livest. Prod. Sci. 72:157-167.

Kok, A., A.T.M. van Knegsel, C.E. van Middelaar, H. Hogeveen, B. Kemp and I.J.M. de Boer, 2015. Technical note: Validation of sensor-recorded lying bouts in lactating dairy cows using a 2-sensor approach. J. Dairy Sci. 98:7911-7916.

Papst, F., O.Saukh, K. Roemer, F. Grandl, J. Jakovljevic, F. Steininger, M. Mayerhofer, J. Duda and C. Egger-Danner, 2019. Embracing opportunities of livestock big data integration with privacy constraints. Proceedings of the 9th International Conference on the Internet of Things (IoT). October 2019. 27:1-4.

Papst, F., K. Schodl and O. Saukh, 2021. Exploring co-dependency of IoT data quality and model robustness in precision cattle farming. SensSys '21: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems. November 2021:433-138.

Paulheim, H. and R. Meusel, 2015. A decomposition of the outlier detection problem into a set of supervised learning methods. Mach. Learn. 100:50-531.

Reith, S., H. Brandt and S. Hoy, 2014. Simultaneous analysis of activity and rumination time, based on collar-mounted sensor technology, of dairy cows over the per-estrus period. Livest. Sci. 170:219-227.

Stygar, A., Y. Gómez, G.V. Berteselli, E. Dalla Costa, E. Canali, J.K. Niemi, P. Llonch and M. Pastell, 2021. A systematic review on commercially available and validated sensor technologies for welfare assessment of dairy cattle. Fron.t Vet. Sci. 8:634338.

Using precision technology to detect the onset of digital dermatitis in dairy cows

J. Magana and A.L. Adams Progar

Washington State University, P. O. Box 646310, 99163, Pullman, USA
Corresponding Author: amber.adams-progar@wsu.edu

Abstract

Digital dermatitis (DD) is a multifactorial infectious disease of the hoof that causes inflammation and painful lesions primarily located digitally and on the coronary band. Changes in cow behaviors are associated with lameness, yet behavioral data related to the onset of DD is limited. The aim of this study was to evaluate behavioral differences between cows with healthy feet and cows with DD, as well as changes in behavior associated with the onset of DD. Lactating Holstein cows (n = 42) were observed in the parlor daily June-July 2020, for visual hoof evaluations. Behavior data was collected for 1 wk prior to the onset of DD using CowManager® activity monitoring ear tags that recorded activity, eating, and ruminating behaviors, and ear temperature. Data were analyzed using mixed model ANOVA and linear regression in SAS. An interaction was detected between day relevant to diagnosis of DD (Pre-diagnosis, 0/Day of diagnosis, Post-diagnosis) and hoof health (healthy or DD) for active ($P < 0.0001$) and eating ($P < 0.0001$) behaviors. Cows with DD tended ($P = 0.08$) to spend 0.07 ± 0.01 h/d less time active prior to a DD diagnosis. A three-way interaction was observed among day relative to the onset of DD, hoof health, and cow lactation number for inactivity ($P = 0.03$) and high activity ($P = 0.05$). Prior to the onset of DD, cows spent 0.03 ± 0.01 h/d less time highly active ($P = 0.04$). Ear temperature was associated with day relative to the onset of DD and hoof health ($P < 0.0001$). Prior to the onset of DD, average ear temperature increased by 0.16 ± 0.03°C ($P < 0.0001$). In conclusion, cows that developed DD altered their behaviors prior to diagnosis. Understanding the progression of this disease could promote early treatment and better prognosis.

Keywords: Cow behavior, digital dermatitis, ear tag sensors.

Introduction

Digital dermatitis (DD) is a multifactorial infectious disease that causes inflammation and is classified by painful lesions primarily located digitally and on the coronary band of the hoof. DD is typically accompanied by lameness, due to the lesions, along with other infectious and non-infectious diseases including: foot rot; sole ulcers; sole hemorrhages; and white line disease. Digital Dermatitis severity can be assessed on a scale of M0 through M4 stages depending on the size, color, and pain of the lesion (Döpfer *et al.*, 1997). The accompanying symptom of digital dermatitis, lameness, negatively impacts animal welfare in dairy cattle and leads to economic loss in the dairy industry (USD$1.1 billion; Zinicola *et al.*, 2015). DD affects 25% of dairy farms and has led to a large economic loss partially due to a decrease in milk production (0.6 kg/day; Relun *et al.*, 2013).

Literature surrounding digital dermatitis remains limited and has been encompassed under the condition lameness. Lameness impacts changes in animal behavior such as lying time which in turn can cause decreases in the amount of time spent eating and feed intake, leading to reductions in milk production (Yunta, 2012; Norring *et al.*, 2014; Argaez-Rodriguez, 1998). Although behavioral data such as these have been studied in lame cows, behavioral data related to the detection of digital dermatitis in cows remains very limited. Behavioral monitoring systems range from neck collars to detect head movements and assess eating and ruminating time, while ankle bracelets can be used to detect pedometry and assess activity time. CowManager® (CowManager B.V., Harmelen, Netherlands) is a common behavioral monitoring system on dairy farms. CowManager® receives its data using ear tags that record daily behaviors such as rumination, eating time, ear temperature, and activity time. These systems have been used as reliable methods to compare an individual to the herd population and detect lameness as well as other diseases in cattle.

Currently, the most common method of DD detection is observational through changes in gait assessed through locomotion scoring, and physical examination of the hoof (Harris-Bridge *et al.*, 2018; Mohamadnia and Khaghani, 2013). Studies using precision technology, such as behavioral monitoring systems, to monitor DD has primarily focused on detection (the mere presence of DD), yet little has been done to evaluate the onset of DD using changes in animal behavior. Due to the progression of the disease, it is not usually noticed until it causes pain that results in cattle putting minimal pressure on the hoof, leading to lameness. The inability to detect the onset of DD impacts animal well-being and is a vital area of research that should be investigated to promote early treatment and a better prognosis.

Materials and methods

All procedures used in this study were approved by the WSU IACUC (ASAF# 6770). The study was conducted from June 2020 to July 2020 at the Washington State University Knott Dairy Center (KDC) in Pullman, Washington, USA. The facility houses 180 lactating Holstein cows in free-stall barns. As a prevention method for DD, an acidified copper sulfate and zinc footbath solution was placed at the exit of the milking parlor. The footbath solution was replaced twice a week through recommendation of a hoof specialist.

Visual assessment of hoof health

The lead researcher for this study was trained by a hoof specialist to evaluate digital dermatitis (DD) lesions. All hoof assessments were conducted during morning milkings in the milking parlor from 0900 h – 1300 h. Lesion status was categorized as active (red and painful with hair on lesion) or degressing (no hair or little hair that lies flat, no pain, and scabbing on lesion). Lesion size was categorized as small (< 0.635 cm), medium (0.635 cm <x< 3.81 cm), and large (> 3.81 cm) by diameter. Cattle were enrolled into the study if they were healthy (no hoof lesions) for at least 7 days before the first observation of an active lesion. Each cow with DD was matched with a healthy counterpart that had the same lactation number, reproduction status, and similar days in milk. The same researcher recorded DD status and size daily until the end of the study.

Cow behavior evaluation

Each lactating cow at the Knott Dairy Center wore a CowManager® (CowManager B.V., Harmelen, Netherlands) ear tag that continuously recorded each cow's behavior 24 hours every day throughout the study. The behaviors of interest in this study were activity, eating, and lying. Once a cow was enrolled into the study, the behavioral data was matched with a healthy cohort according to lactation number, days in milk (or lactation period), and reproductive status. All behavioral data were calculated as the proportion of time each cow spent exhibiting each behavior every day.

Statistical analysis

Prior to statistical analysis, the dependent variables for behavior data were tested for normality using the UNIVARIATE procedure of SAS (SAS 9.4, SAS Institute, Cary, NC, USA) and no transformation of the data was required.

The analyses were performed using a mixed model ANOVA in SAS (PROC MIXED; SAS 9.4, SAS Institute, Cary, NC, USA) for dependent variables: non-active; active; high active; eating; ruminating; and ear temperature. Day in relation to the detection of DD was divided between pre-DD, day 0, and post-DD. Statistical significance was classified as P d" 0.05. Cows that developed digital dermatitis during the study (n = 21) were compared to healthy cows (n = 21).

Cows enrolled into the onset cohort were required to meet the following criteria: > 7 days healthy and > 2 consecutive days of DD lesion present. Statistical analysis for the onset of DD dataset (n = 18) was performed using a regression model in SAS (SAS 9.4, SAS Institute, Cary, NC, USA) for dependent variables: non-active, active, high active, eating, ruminating, and ear temperature. The independent variables used for this model included day in relation to detection of DD. Day 0 was relative to first day of diagnosis and days prior followed a -1, -2, -3, -4….. day sequence until the first day of data collection.

Results

Healthy cows versus cows with DD

Non-active behavior

A three-way interaction was detected when testing non-active behavior between hoof health (DD or healthy), lactation number, and day relative to diagnosis of DD (P = 0.03; Figure 1).

When comparing cows with DD and healthy cows within day 0 of diagnosis, first lactation cows with DD spent 1.76 ± 0.74 h/day less time non-active than healthy cows within their third or higher lactation (P = 0.02). This was also seen pre-diagnosis of DD where healthy cows with three or more lactations exhibited 1.64 ± 0.69 h/day more non-active behavior compared to first lactation cows with DD (P = 0.02).

Healthy cows demonstrated changes in non-active behavior within stages of lactation. As healthy cows increased in number of lactations, non-active time increased (Figure 2). This pattern was observed in all three stages relevant to day of diagnosis (pre, 0, post). Most significantly, when comparing healthy cows at the beginning of the study to the cows closer to the end of the study, those with lower lactations (Lact 1) had lower non-active times compared to higher lactation cows (Lact 3+) (P = 0.006). Fewer differences were found within each category relevant to day of diagnosis, yet there was a trend at the beginning of the study (group pre) between lactation 1 and cows with three or more lactations (P = 0.05).

Cows with DD showed greater variation in non-active behavior when comparing the effect of lactation number. Prior to day of diagnosis (pre), cows with higher lactation numbers exhibited more non-active behavior. Cows with one lactation had a mean non-active time of 6.69 ± 0.47 h/day while cows with three or more lactations had a mean of 7.63 ± 0.51 h/day. Upon diagnosis on day 0, cows with two lactations had more time spent non-active than those with one lactation (P = 0.04). Cows with two lactations spent 1.39 ± 0.66 h/day more mean non-active time than those with one lactation (P

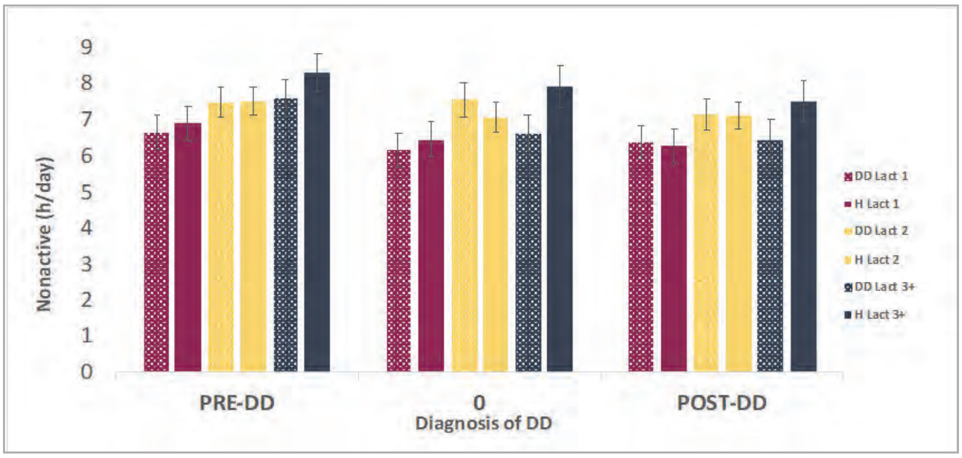


*Figure 1. Mean non-active behavior for cows with DD (n = 21) and healthy (n = 21) cows with different lactation numbers and grouped by period relative to diagnosis (pre-diagnosis, 0/day of diagnosis, and post-diagnosis). * P = 0.02.*

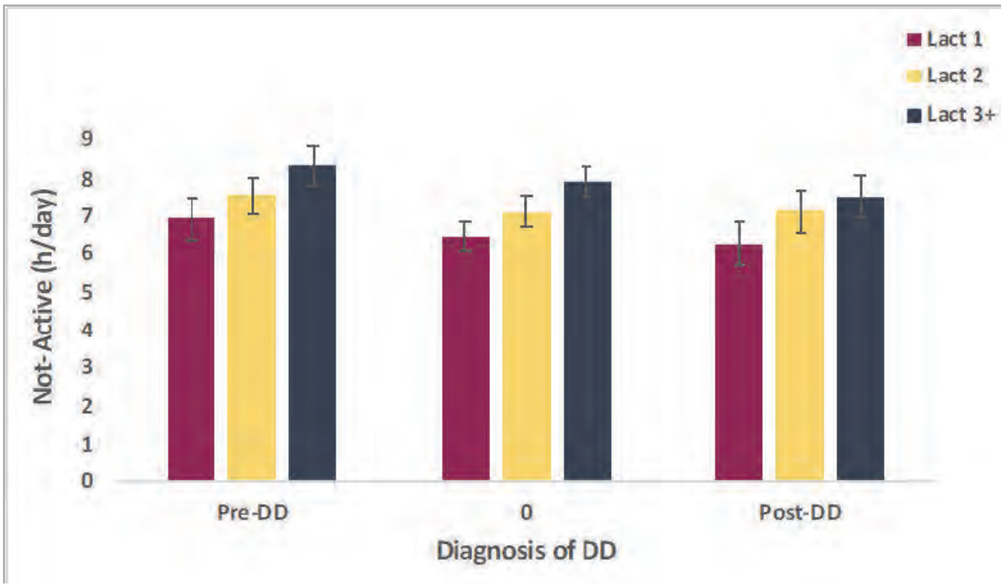


Figure 2. Mean non-active behavior for healthy cows (n=21) and DD cows (n=21) with different lactation numbers and grouped by period relative to diagnosis (pre-diagnosis, 0/ day of diagnosis, and post-diagnosis).

= 0.04). Cows with three or more lactations exhibited 0.96±0.70 h/day less non-active behavior than those in lactation two. Similarly, cows with two lactations spent more time non-active than cows with one or three or more lactations post-DD diagnosis. When comparing non-active behavior among cows with DD, cows with three or more lactations pre-diagnosis of DD had less non-active time than post-diagnosis (P = 0.005).

High active behavior

A three-way interaction was detected for high active behavior among hoof health, lactation number, and day relative to DD diagnosis (P = 0.05; Figure 3).

Healthy cows demonstrated changes in high active behavior at different lactation numbers. Generally, cows with one lactation had greater than or equal to high-active time compared to cows with two and three or more lactations with the highest mean of 3.04 ± 0.31 h/day. When assessing differences in DD cows pre-diagnosis of DD, differences were noticed between cows with one lactation and three or more lactations (P = 0.02). When comparing cows within group day 0, DD cows with one lactation had an increase in mean high active time of 1.37 ± 0.46 h/day compared to cows with three or more lactations (P = 0.004). DD cows with one lactation post day of diagnosis also showed higher high active time compared to those with two lactations (P = 0.01) and three or more lactations (P = 0.03). Not many differences were seen within stages of lactation with the exception of DD cows with two lactations. Comparison between 0 day of diagnosis and post day of diagnosis yielded an increase of high active time by a mean of 0.37±0.15 h/day (P = 0.02). Fewer differences were noticed when comparing healthy and DD cows within pre-DD, 0, and post-DD. Upon day of diagnosis a significant increase in high-active behavior was shown amongst first lactation cows (P = 0.02).

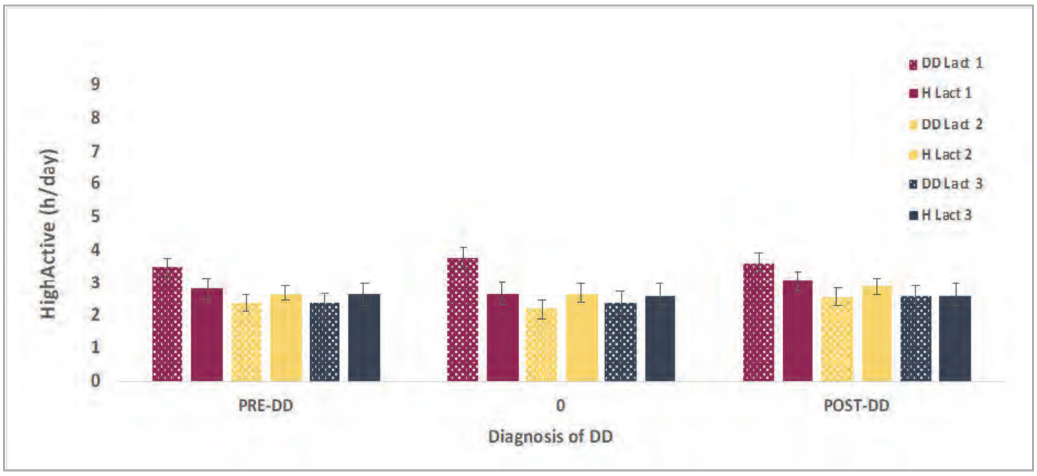


Figure 3. Mean high active behavior for cows with DD (n = 21) and healthy (n = 21) cows with different lactation numbers and grouped by period relative to diagnosis (pre-diagnosis, 0/day of diagnosis, and post-diagnosis).

Eating behavior

A two-way interaction among day relevant to diagnosis of DD and hoof health was detected for eating behavior (P < 0.0001). Cows that were healthy showed highest eating behavior with a mean eating time of 3.59±0.24 h/day. DD cows showed increase in eating behavior as progression of DD occurred. At pre-DD, cows had a mean eating time of 3.13±0.24 h/day and increased upon day 0 of DD to 3.47±0.25 h/day (P = 0.004). Post-DD eating behavior increased by 0.03±0.11 h/day, yet was not significant. When comparing DD cows at the beginning of the study within group pre-DD and post-DD, eating behavior increased by a mean of 0.37±0.11 h/day (P = 0.0009).

Active behavior

A two-way interaction was observed between hoof health and day relevant to DD diagnosis for active behavior (P < 0.0001). Within pre-DD, day 0, and post-DD, no significant differences were shown between healthy and DD cows, yet differences are seen between day of DD and DD cows. Pre-DD cows showed 0.23±0.08 h/day less mean active time than on day 0 (P = 0.004). After initial diagnosis on day 0, DD cows slightly increased in mean active time (0.002±0.08 h/day) but was not significant. When comparing pre-DD cows to post-DD cows, differences indicated an increase of 0.23±0.07 h/day (P = 0.003). When comparing healthy and DD cows at different days of diagnosis, healthy cows had lower mean active time then DD cows at day 0 of DD. Mean active time increased by 0.28±0.12 h/day upon diagnosis of DD (P = 0.03). Comparing healthy cows at the beginning of the study and DD cows at the end of the study (post-DD) showed an increase of 0.28±0.12 h/day in mean active time (P = 0.02).

Onset of DD

High active behavior

Linear regression model indicated a difference in high active behavior and day relative to diagnosis of DD (P = 0.04). When comparing high active behavior, a slight negative relationship was determined as cows approached initial day of DD. As cows neared the day of DD diagnosis, they spent 2.58 h/day less time being high active. Although there was a significant relationship between high active behavior and day, R^2 (0.023) remained low.

Eating behavior

Linear regression model indicated a difference in eating behavior and day relative to diagnosis of DD (P = 0.002). As cows neared the day of DD diagnosis, they spent 3.30 h/day more time eating. When comparing eating behavior, a slight positive correlation was determined as cows approached initial day of DD. Although there was a significant correlation between eating behavior and day, R2 (0.0496) remained low.

Discussion

In this study, healthy cows spent an average of 7.25 ± 1.48 h/day being non-active, which is similar to a study that discovered a range of 7.2 to 9.34 h/day of non-active behavior for cows (Bikker *et al.*, 2014). Studies have found healthy cows spend 3.2 to 5.4 h/day eating which is also similar to those included in this study (3.22 ± 1.24 h/day; Gomez and Cook, 2010).

Although ruminating time was not significant in this study, the mean time was within the range of those seen in previous studies (8.62 ± 1.14 h/day; Reynolds *et al.*, 2019). Bikker and colleagues (2014) grouped high-active and active behavior into one behavior category and found that cows spend, on average, 3.58 h/day being active, which is slightly higher in comparison to the average time spent active for cows in this study (2.51 ± 0.80 h/day). Interestingly, healthy cows at the beginning of the study (group Pre) had lower active time compared to those at the end of the study. This study was conducted over the summer and studies have demonstrated that heat stress can cause changes in normal behaviors such as activity time (Zahner *et al.*, 2004; Cook *et al.*, 2007). The use of stored weather data at three different times points within the study demonstrated temperature changes. At the beginning of the study temperatures averaged 19°C, increased half way through (24°C) as well as at the end of the study (30°C)(Accuweather, n.d).

The data in this study also showed variation in inactive and high active time at different stages of lactation in healthy cows. Cows with higher lactation numbers (3+) displayed more inactive behavior relative to day of diagnosis. Overall, cows in lactations 1 and 2 had more time spent high active compared to those with three or more lactations. This behavior is similar to those seen in other articles that tested for the effects of age on normal behaviors. Steensels *et al.* (2012) concluded that lying time was increased with age by 29 min/day in cows with three or more lactations compared to those with only two lactations. In general, studies determined that animals exhibited less active behavior as they become older (Ingram, 2000). In addition, the increase in high-active time of younger cows may be a response from competition and dominance produced from older cows if competing for space. Cows are social animals that will create hierarchies dependent on their ability and willingness to establish dominance. In cases where space may be limited, cows will compete and may put forth aggressive behavior such as bunting and pushing, forcing cows to become more active (Phillips and Rind, 2002; Kondo *et al.*, 1989). An indication of weakness in cows will be taken as an opportunity to challenge each other and achieve a rise in hierarchy.

Weaknesses in cattle can include signs of illnesses such as DD and/or calving where some studies had noted that cows will be weakened by parturition and sudden change of lactation (Lamb, 1976). Cows with DD generally display signs of lameness such as changes in gait that may be noticed by other cows. When comparing activity time in cows with DD and stages of lactation, first lactation cows seemed to be impacted the most upon diagnosis of DD. On Day 0 and Post-DD, lactation 1 cows exhibited less inactive and more high-active behavior then lactations 2 or 3+ cows. Previous studies noted age as an additional point of weakness as they are generally smaller and more timid, making it less likely for them to elicit an aggressive response (Lamb, 1976).

The current findings of healthy and DD cows indicated an impact of disease on normal inactive and eating behaviors. DD cows with three or more lactations showed less inactive behavior compared to their healthy counterpart within Day 0 and Post-DD. DD lesions have been associated with high amounts of pain which could influence these behaviors (Palmer and O'Connel, 2015). Additionally, DD cows showed more eating behavior Post-DD compared to healthy counterparts. These findings differ from those in previous studies that compared lame cows to healthy ones (Weigele *et al.*, 2018). Generally, it is expected that cows with pain would increase their inactive time, spending more time lying compared to healthy cows, yet some studies correlated an increase in active behavior to the increase in shorter strides cows take in order to reduce discomfort (Alsaaod *et al.*, 2012).

Analysis of the onset of DD in cows showed differences in variability of active and eating behaviors in the days leading up to the first day of DD diagnosis (Day 0), but did not for rumination, inactive, and high-active behaviors. On days prior to DD, cows showed more active behavior than on the day of DD diagnosis. This is interesting as it describes a different relationship than the one seen in the previous cohort. The relationship seen in these cows is one that is expected as the development of DD occurs. Previous studies that assessed behaviors in lame cows noted the decrease in active behavior over time (Weigele *et al.*, 2018; Van Nuffel *et al.*, 2015). Mazrier *et al.* (2006) used pedometry in an attempt to detect the onset of lameness in dairy cows and found that lame cows showed a reduction in pedometry activity 7 to 10 days prior to any clinical signs. Eating behavior maintained a positive correlation as the progression of DD occurred in this study. Cows exhibited less eating time compared to those on Day 0 of DD, which differs from previous studies of lameness (Weigele *et al*., 2018).

Conclusion

The objectives of this study were to determine if precision monitoring systems such as behavioral monitoring systems could provide insight into the detection of digital dermatitis in dairy cows. Ear monitoring systems provided behavioral data that could be used to identify shifts in normal behaviors compared to healthy cows and within the onset DD cohort. When comparing DD cows to healthy ones, some behavioral patterns differed upon day of diagnosis and post-DD diagnosis. Onset DD cows demonstrated changes in behavior prior to initial day of DD. Interestingly, differences in behaviors were seen between different lactation numbers of healthy cows and among DD cows. Overall, the use of precision technology could be beneficial in the detection of digital dermatitis of dairy cows. Future research should assess the behaviors of cows with DD in a longer study. Lameness scoring throughout the entire study could help determine differences between healthy and DD cows as well as changes in the onset of DD cows.

Acknowledgement

Many thanks are extended to Dr. Celina Matuk Sarinana and her staff at the WSU Knott Dairy Center for their assistance with data collection. Mr. Bingham’s expertise in conducting hoof evaluations was an invaluable contribution to this project. Thanks are also extended to Anna Elder for her assistance with data collection. We thank Mr. John Swain and Dr. Craig McConnel for their assistance with experimental design and article editing.

References

Accuweather. Retrieved August 25, 2020, from www.accuweather.com/

Alsaaod, M., C. Romer, J. Kleinmanns, K. Hendriksen, S. R. Meierhofer, L. Plumer and W. Buscher, 2012. Electronic detection of lameness in dairy cows through measuring pedometric activity and lying behavior. Appl. Anim. Behav. Sci. 142: 134-141.

Argáez-Rodríguez, F. de J., D. W. Hird, J. H. de Anda, D. H. Read, and A. Rodríguez-Lainz, 1997. Papillomatous digital dermatitis on a commercial dairy farm in Mexicali, Mexico: incidence and effect on reproduction and milk production. Pre. Vet. Med. 32: 275-286.

Bikker, J., H. van Laar, P. Rump, J. Doorenbos, K. van Meurs, G. Griffioen and J. Dijkstra, 2014. Technical note: Evaluation of an ear-attached movement sensor to record cow feeding behavior and activity. J. Dairy Sci. 97: 2974–2979.

Cook, N. B., R. L. Mentink, T. B. Bennett and K. Burgi, 2007. The effect of heat stress and lameness on time budgets of lactating dairy cows. J. Dairy. Sci. 90: 1674-1682.

Döpfer, D., A.A.H.M. ter Huurne, J.L. Cornelisse, A.J.A.M. van Asten, A. Koopmans, F.A. Meijer, Y. H. Schukken, I. Szakáll, W. Klee and R.B. Bosma, 1997. Histological and bacteriological evaluation of digital dermatitis in cattle, with special reference to spirochaetes and Campylobacter faecalis. Vet. Rec. 140: 620-623.

Gomez, A. and N.B. Cook, 2010. Time budgets of lactating dairy cattle in commercial freestall herds. J. Dairy Sci. 93: 5772-5781.

Harris-Bridge, G., L. Young, I. Handel, M. Farish, C. Mason, M. A. Mitchel and M. J. Haskell, 2018. The use of infrared thermography for detecting digital dermatitis in dairy cattle: What is the best measure of temperature and foot location to use? Vet. J. 237: 26-33.

Ingram, D.K, 2000. Age-related decline in physical activity: generalization to nonhumans. Med. Sci. Sports. Exerc. 32: 1623-1629.

Kondo, S., J. Sekine, M. Okubo and Y. Asahida, 1989. The effect of group size and space allowance on the agonistic and spacing behavior of cattle. App. Anim. Behavi. Sci. 24: 127-135.

Lamb, R. C. 1976. Relationship between cow behavior patterns and management systems to reduce stress. J. Dairy. Sci. 59: 1630-1636.

Mazrier, H., S. Tal, E. Aizinbud and U. Bargai, 2006. A field investigation of the use of the pedometer for the early detection of lameness in cattle. Can. Vet. J. 47: 883-886.

Mohamadnia, A. and A. Khaghani., 2013. Evaluation of hooves' morphometric parameters in different hoof trimming times in dairy cows. Vet. Res. Forum. 4: 245-249. PMCID: PMC4279619.

Norring, M., J. Häggman, H. Simojoki, P. Tamminen, C. Winckler, and M. Pastell, 2014. Short communication: lameness impairs feeding behavior of dairy cows. J Dairy Sci. 97: 4317-4321.

Palmer, M. A. and N. E. O'Connell, 2015. Digital dermatitis in dairy cows: a review of risk factors and potential sources of between-animal variation in susceptibility. Anim. 5: 512–535.

Phillips, C. J. C. and M. I. Rind, 2002. The effect of social dominance on the production and behavior of grazing dairy cows offered forage supplements. J. Dairy. Sci. 85: 51-59.

Relun, A., A. Lehebel, A. Chesnin, R. Guatteo and N. Bareille, 2013. Association between digital dermatitis lesions and test-day milk yield of Holstein cows from 41 French dairy farms. J. Dairy. Sci. 96: 2190-2200.

Reynolds, M. A, M. R. Borchers, J. A. Davidson, C. M. Bradley and J. M. Bewley. Technical note: An evaluation of technology-recorded rumination and feeding behaviors in dairy heifers. J. Dairy Sci. 102: 6555-6558.

Steensels, M. C. Bahr, D. Berckmans, I. Halachmi, A. Antler and E. Maltz. Lying patterns of high producing healthy dairy cows after calving in commercial herds as affected by environmental condition and production. App. Anim. Behav. Sci. 136: 88-95.

Van Nuffel, A., I. Zwertvaegher, L. Pluym, S. Van Weyenberg, V. M. Thorup, M. Pastell, B. Sonck and W. Saeys, 2015. Lameness detection in dairy cows: part 1. how to distinguish between non-lame and lame cows based on differences in locomotion or behavior. Anim. 5: 838-860.

Weigele, H., C. L. Gygax, A. Steiner and J. B. Burla, 2018. Moderate lameness leads to marked behavioral changes in dairy cows. J. Dairy. Sci. 101: 2370-2382.

Yunta, C., I. Guasch, and A. Bach, 2012. Short communication: lying behavior of lactating dairy cows is influenced by lameness especially around feeding time. J. Dairy Sci. 95: 6546-6549.

Zahner, M., L. Schrader, R. Hauser, M. Keck, W. Landhans and B. Wechsler, 2004. The influence of climatic conditions on physiological and behavioural parameter in dairy cows kept in pen stables. Anim. Sci. 78: 139-147.

Zinicola, M., F. Lima, S. Lima, V. Machado, M. Gomez, D. Döpfer and R. Bicalho, 2015. Altered microbiomes in bovine digital dermatitis lesions, and the gut as a pathogen reservoir. PLOS. ONE. 10: 2-23.

Milk recording in sheep and goat: state of the art using the data from the ICAR on-line yearly survey

J.M. Astruc[1], A. Carta[2], R. Negrini[3], M. Simcic[4], M. Spehar[5], E. Ugarte[6] and C. Mosconi[7]

[1]*IDELE, Castanet-Tolosan, France.*
Corresponding Author: Jean-Michel.Astruc@idele.fr
[2]AGRIS Sardegna, Sassari, Italy
[3]AIA, Roma, Italy
[4]University of Ljubljana, Ljubljana, Slovenia
[5]Croatian Agency for Agriculture and Food, Zagreb, Croatia
[6]NEIKER, Vitoria, Spain
[7]ICAR, Utrecht, The Netherlands

Abstract

From the nineties, the working groups in charge of small ruminants in ICAR carried out surveys on topics related to milk recording and breeding programs in sheep and goats. These surveys were manual from 1988 to 2006, before being replaced by a yearly on-line questionnaire available to accept submission of data from ICAR countries having an activity in the field of dairy sheep and goats recording. This paper aims to analyse the evolution of milk recording in sheep and goats over the years, using the database (developed and maintained within the umbrella of ICAR Secretariat) updated by the responses from the on-line surveys. The database includes data on milk recording activities (by countries and by breeds), milk recording designs, average milk yield per lactation, recording devices, breeding schemes and selection criteria, and molecular information in sheep and goats. The valorisation of the on-line survey constitutes one of the main terms of reference of the sheep, goats and camelids working group and a synthesis is regularly presented at the working groups meeting. An overlook of the main evolutions of the milk recording activities in sheep and goats are presented, over the last 35 years. Ten to twelve countries filled out the questionnaire in each species. The number of animals in official milk recording reached 890,000 ewes and 410,000 does in 2021. In the large populations of Mediterranean countries, the impact of qualitative recording remains low, due to its cost. Use of simplified designs, which has been recommended by the working group has strongly increased over the years to exceed 90%.

Keywords: Dairy sheep, dairy goats, milk recording, recommendations.

Introduction

One of the objectives of the terms of reference of the Sheep, Goat and Camelid Working Group (SGC-WG) of ICAR is to conduct and report the results of periodic international surveys on sheep, goat and camelid performance recording and genetic evaluation. In its history, the SGC-WG was created in two steps. The first step was the creation in 2016 of the sheep, goat and small camelid working group from three existing working groups (WG): the Performance Recording of Dairy Sheep WG created in 1988, the Goat Performance Recording WG and the Animal Fiber WG (dedicated to mohair goats and small camelids). With this first step, meat sheep and goats were included in the scope of the new working group. The second step occurred in 2020 with the inclusion of the dromedaries and Bactrian camels.

The first survey in dairy sheep was carried out “manually” in 1988 and was followed by several manual surveys every 2 years until 2004. Meanwhile, the same type of survey was conducted in goats between the nineties and 2006. From 2006 (dairy sheep) and 2008 (goat), an on-line database with on-line submission of data was set up at the ICAR secretariat level. This survey could be filled in on a yearly basis. Since 2021, a new platform for collecting data was implemented, permitting a more friendly submission of data and allowing easy extractions and statistics for ICAR members.

So far, no such questionnaire has been implemented neither in meat sheep and goats, nor in camelids.

This paper aims at proposing a synthesis of the main results of the database, by giving a current vision of the situation of milk recording, while also giving an overview of the evolution of milk recording over the last decades in small ruminants.

The ICAR on-line survey

The ICAR on-line survey follows the same pattern in dairy sheep and goats, with the same sections and the same data to submit. The enquiry is divided in 9 sections, representing 7 different items.

- Basic information on population, milk recording and management of the lactation.
- Methods of milk recording.
- Type of lactation calculation for milk yield.
- Milk yield results.
- Optional test for milk composition.
- Recording of non-milking traits.
- Milk recording equipment used in case of machine milking
- Breeding programs using artificial insemination (AI).
- Molecular information.

The objective of the survey is to have a state of the art of the situation of milk recording in ICAR countries. It also allows to follow the different recommendations of the working group (use of simplified recording method, type of lactation calculation) as they are mentioned in the ICAR guidelines.

Few countries filled the questionnaire. Over the last 2 years, 7 countries filled out at least once in sheep (France, Italy, Spain, Slovak Republic, Czech Republic, Slovenia, Croatia) and 9 countries filled out at least once in goat (France, Italy, Spain, Slovak Republic, Czech Republic, Slovenia, Croatia, Serbia, Latvia). In addition, between 10 and 2 years ago, 4 countries submitted data in sheep (Canada, Germany, Greece, Portugal) and 4 countries in goat (Canada, Germany, Switzerland, Portugal).

We will specifically focus in this paper on the following topics: sheep and goats in milk recording, recorded breeds, recording methods, lactation calculation methods.

On the whole, if we aggregate data from all the countries that have submitted data for the last 10 years, 877,896 sheep and 457,324 goats were recorded. 86% and

Recorded sheep and goats in milk recording and evolution over the last decades

88% (respectively in sheep and goats) were recorded in the following three countries: France, Spain and Italy. The percentage of recorded females is 8.5% in sheep and 9.6% in goats, which is much less than in dairy cattle. The detail by country is given in the table 1 (dairy sheep) and 2 (dairy goat).

The evolution of the number of recorded females varies across countries, as shown in the figures 1, 2, 3 and 4. In dairy sheep, the trend is quite stable in France, increasing in Spain. In Italy, after a growing period between 1988 and 2012, the number of recorded ewes dramatically decreased over the last 10 years and has been divided by 3. The evolution is stable in Czech, Slovenia and Croatia, decreasing in Slovak.

In goat, the depth of history is less important: the number of recorded females is stable in Spain, Italy, Latvia, Slovak, Czech, slightly decreasing in France, increasing in Slovenia, Croatia and Serbia over the last 10-15 years.

Table 1. Size of population of dairy sheep, impact of quantitative and qualitative recording in ICAR member countries (2020-2021).

			Qualitative recording	
Countries	Population size (*)	Quantitative recording. Number of recorded does (official milk recording)	Yes/Not	Number of recorded ewes
Croatia	95,000	7,235	Yes	7,020 (~100%)
Czech Republic		1,494	Yes	All
France	1,604,000	334,685 (**)	Yes	104,606 (31%)
Italy	4,851,000	161,711	Yes	10,895 (7%)
Portugal	263,000	18,052	No	-
Slovak Republic	163,000	6,643	Yes	6,643 (100%)
Slovenia	5,000	4,624	Yes	4,624 (100%)

Table 2. Size of population of dairy goats, impact of quantitative and qualitative recording in ICAR member countries (2020-2021).

			Qualitative recording	
Countries	Population size (*)	Quantitative recording. Number of recorded does (official milk recording)	Yes/Not	Type of recorded does
Croatia	66,000	3,621	Yes	All
Czech Republic		5,152	Yes	All
France	1,193,000	227,955	Yes	All
Germany		1,517	Yes	All
Italy	826,000	60,326	Yes	All
Latvia	7,900	1,296	Yes	All
Portugal	247,000	7,771	Yes	-
Serbia	119,000	4,846		82%
Slovak Republic	33,000	346	Yes	Parities 1-3
Slovenia	4,800	2,575	Yes	All
Spain	1,890,000	113,934	Yes	-

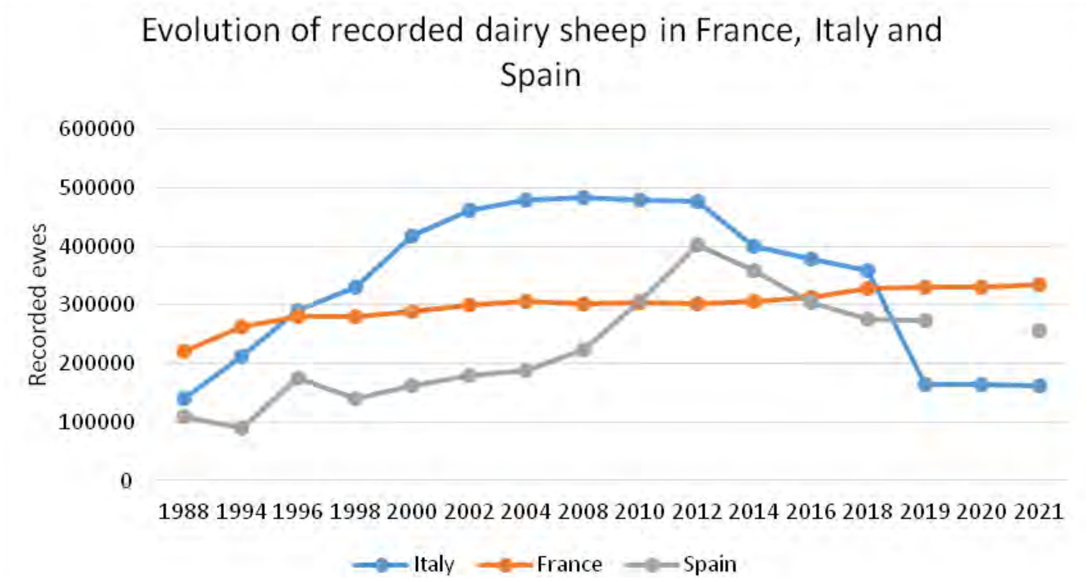


Figure 1. Evolution of recorded dairy sheep in France, Italy and Spain since 1988.

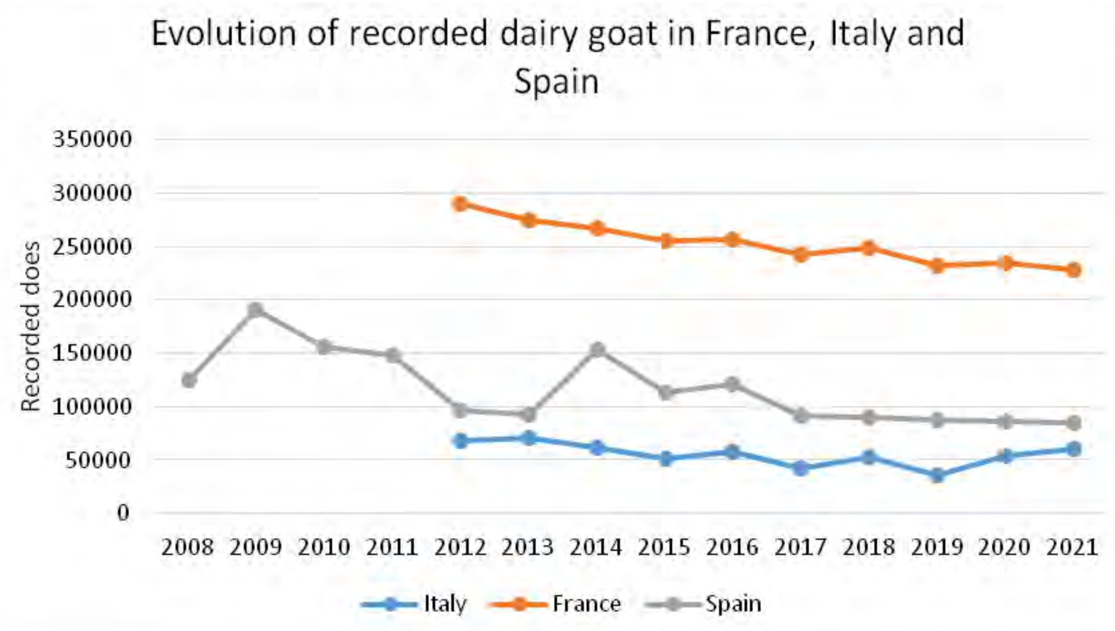


Figure 2. Evolution of recorded goats in France, Italy and Spain since 2008.

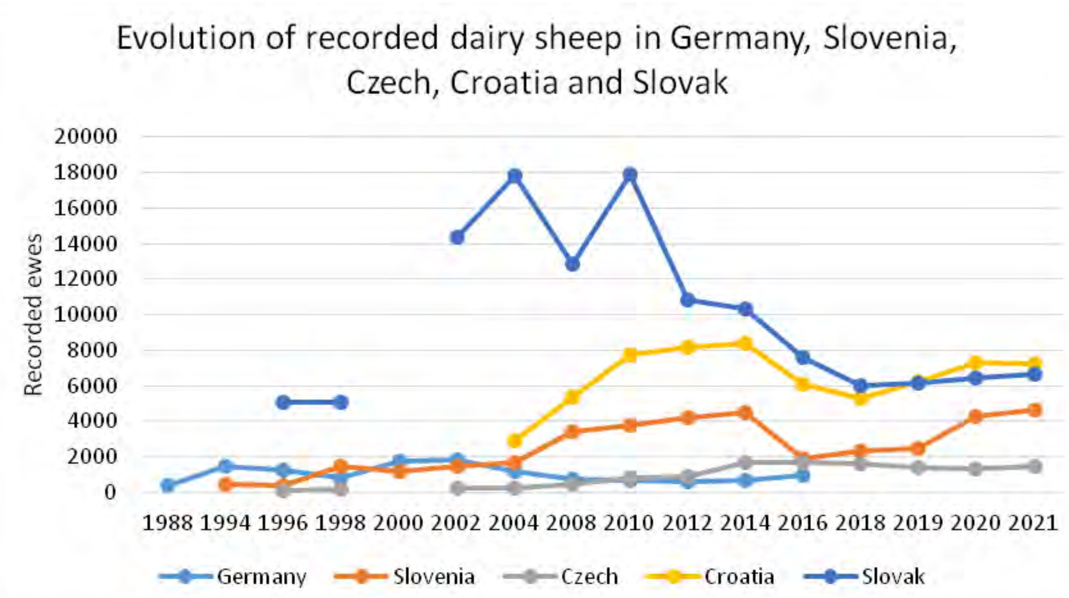


Figure 3. Evolution of recorded dairy sheep in Czech, Slovak, Slovenia, Croatia, Germany since 1988.

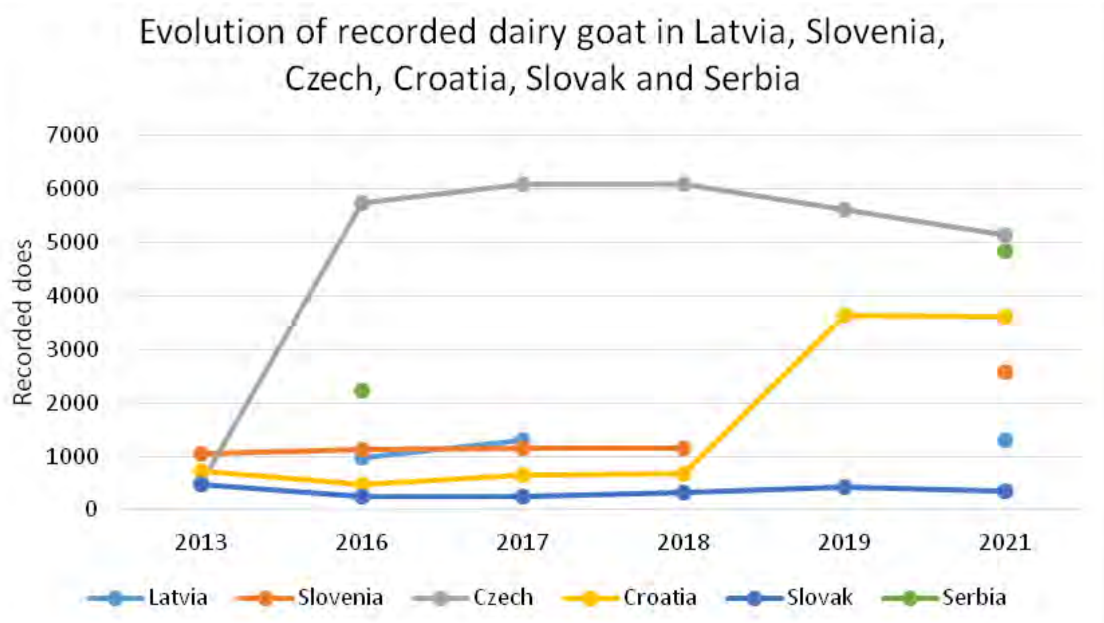


Figure 4. Evolution of recorded goats in Czech, Slovak, Slovenia, Croatia, Serbia, Latvia since 2013.

Table 3. Main recorded breeds in sheep and goat in ICAR member countries (2020-2021).

Species	Breed	Country	Number of recorded females
Goat	Alpine	France	146,854
	Saanen	France	80,750
	Murciano-Granadino	Spain	67,543
	Malaguena	Spain	12,863
	Florida	Spain	14,676
	Saanen	Italy	11,606
	Camosciata delle Alpi	Italy	13,573
Sheep	Lacaune	France	192,900
	Assaf	Spain	141,000
	Manchega	Spain	136,000
	Sarda	Italy	115,500
	Manech tête rousse	France	84,300
	Latxa	Spain	72,000
	Churra	Spain	26,000

A great variety of local breeds, few international breeds

The description by breeds shows a variety of local breeds: 40 goat and 35 dairy sheep breeds have been reported in only one country. Besides all these local breeds, some breeds have an international dissemination: in goat, the number of countries in which the breed is recorded reached 6 in Alpine, 5 in Saanen, 2 in Murciano-Granadino and Malagueña. In sheep, this number is 6 (Lacaune), and 3 (Assaf and East Friesian).

The table 3 shows the number of recorded females in the most important breeds within country. Alpine in France is the most recorded goat breed, while Lacaune in France is the most recorded dairy sheep breed.

Recording method and lactation calculation

The table 4 describes the recording designs used in the different countries, as well as the types of lactation calculation that are set up.

Recording methods

As milk recording is more time-consuming and more expensive in sheep and goat than in cattle (size of the flocks/herds, rapid milking routine), the working group has recommended to implement simplified method of recording. Current simplified methods are AT (alternated one-milking recording), AC (one recorded milking per day) while current non simplified method is A4 (or B4) and their derivatives for sampling (AY, AZ, CY, CZ are on the basis of A4 methods, but with 1 sample per day). Simplified methods are quite generalised in sheep (almost 99% in AT or AC), less used in goat (58% in AT or AC). In France, the commercial flocks that are not in selection but want to get a service of milk recording practice the B method which is very flexible and does not result in EBV calculation. In Italy, in the Sarda breed, the procedure of AC method quality control based on one A4 recording along the milking period (as described in the ICAR guidelines) started in 2021.

For qualitative recording, while more or less all the goats submitted to milk recording are sampled, in dairy sheep the situation is more contrasted. Central Europe and Balkan countries (with a small number of recorded ewes) sample all the ewes, while Spain, France, Italy only sample a part of the ewes (part of the flocks in qualitative recording; part of the ewes within flock, according to some parities). The ratio of ewes submitted

Table 4. Recording methods and expression of milk yield in sheep and goat in ICAR member countries (2020-2021).

			Type of lactation calculation	
Countries	**Species**	**Recording methods**	**Lactation**	**Production of reference**
Croatia	Goat	AT(most) - A4	TSMM,TMM	
	Sheep	AT(most) – B4	TSMM,TMM	
Czech Rep.	Goat	AC - E		
	Sheep	AT		
France	Goat	A4,AY,AZ,CY,CZ – AT - AC	TMY	
	Sheep	AC - B	TMM	
Italy	Goat	AT	TSMM,TMM	TSMM, TMM
	Sheep	AT – AC (Sarda)	TSMM,TMM	TMM
Latvia	Goat	A4	TMY	TMY (350)
Portugal	Goat	A4(most) - AT	TMM	TMM (90-120)
	Sheep	A4(most) - AT	TMM	TMM, TMY (150)
Serbia	Goat	AT	TSMM,TMM,TMY	TSMM,TMM,TMY
Slovak Rep.	Goat	AC	TMM	TMM (240)
	Sheep	AC	TMM	TMM (150)
Slovenia	Goat	AT	TSMM,TMM,TMY	
	Sheep	AT	TSMM,TMM,TMY	
Spain	Goat	A4 - AT - AC	TMM,TMY,TSMM	TMM (150 or 254), TMY (210 or 274), TSMM (210)
	Sheep	AT – AC (Latxa-part)	TMM,TMY,TSMM	TSMM (116 or 120), TMY (163), TMM (120)

to sampling is 31% in France, 7% in Italy, and is estimated to hardly above 50% in Spain. In addition, it exists protocols to sample only a part of the lactation, either the most representative part on a genetic point of view (part-lactation sampling within AC method) or one out of two successive test-days (case of AT method in Spanish Manchega breed).

Lactation calculation

The definition of the lactation calculation derives from the exploitation of the lactation in sheep and goat. There are 2 main systems, according to the presence or absence of a suckling period where the lambs/kids remain with their dam and where there is no milking or a mixed milking plus suckling. In goats, the main system is without suckling period, at least in the international Alpine and Saanen breeds. But the system with a suckling period is still important in local, less productive breeds. In sheep, in most breeding systems, the lactation is divided into two parts: a suckling period (of about 1 to 2 months) either without any milking or with a partial milking - once a day for example - and a milking-only period after the weaning of the lamb(s).

When there is no suckling period, the lactation is de facto calculated from the lambing: this is the Total Milk Yield (TMY). When there is a suckling period, there are two kinds of calculation: the Total Milked Milk (TMM) or the Total plus Suckled Milk Milked (TSMM). Besides these methods, it is possible to calculate the lactation on the actual lactation length or on a reference length.

The table 4 shows the diversity of situation across country and sometimes breeds, rendering impossible the comparison of the production of the different breeds. Some countries have one type of calculation (case of France or Slovak). Slovak calculates on the same way the lactation of sheep and goats, whereas France calculates TMY in goats and TMM in sheep. Spain has a large variety of calculation, across breeds, and also within breed.

Conclusion and perspectives

On-line survey is a relevant tool to follow the state of the art in dairy sheep and goat milk recording and related topics. However, countries with a large number of small ruminants are missing (Greece, some of Balkan countries). These countries should be stimulated and encouraged to submit data to the on-line questionnaire. Milk recording in sheep and goat represents a small proportion of the populations: 8.5% in sheep and 9.6% in goat. Simplified methods (recording, sampling) are key to increase milk recording. Finally, we suggest that such a survey should be proposed in meat sheep/goat and in wool recording in the future. It might be an objective of the SGC WG.

References

FAOSTAT, 2020 - http://www.fao.org/faostat/en/#data/QCL

ICAR guidelines – Section 16 Dairy Sheep and Goats. https://www.icar.org/Guidelines/16-Dairy-Sheep-and-Goats.pdf

ICAR on-line database for cow, sheep and goat milk recording. https://my.icar.org/

Wool recording in sheep: results from an ICAR on-line survey

*M. Spehar[1], M. Antonini[2], J. Conington[3], S. McIntyre[4], D. Brown[5], M. Simcič[6],
K. McDermott[7], C. Mosconi[8] and J.M. Astruc[9]*

*[1]Croatian Agency for Agriculture and Food, Zagreb, Croatia
Corresponding Author: marija.spehar@hapih.hr
[2]ENEA, Italy
[3]SRUC, United Kingdom
[4]Beef and Lamb NZ Genetics, New Zealand
[5]Animal Genetics and Breeding Unit, Australia
[6]University of Ljubljana, Ljubljana, Slovenia
[7]Sheep Ireland, Ireland
[8]ICAR, Utrecht, the Netherlands
[9]IDELE, Castanet-Tolosan, France*

Abstract

The ICAR Sheep, Goats and Camelids Working Group based on the interests of breeders' organisations has decided to include wool performance recording in its guidelines. Therefore, the Working Group has established an Expert Advisory Group with the objectives to determine the traits of interest, their collection and genetic evaluation. In order to achieve these objectives, representatives of breeding organisations and other relevant institutions responsible for sheep recording were asked to fill out an on-line survey on sheep wool. The survey includes basic information such as contact details of the organisations and specific questions regarding the wool population. The latter includes information's of the breeds involved, the size of the recorded and non-recorded population, the number of farms with wool records, information about traits recorded, phenotyping methods and protocols, genetic evaluation, and selection indices. Additional relevant data on management of animals (e.g., shearing) were also collected in the survey. Altogether, seventeen breeding organisations responded to the on-line survey. The number of animals included in the wool recording by country ranged from 200 to 2,900,000. As expected, the most common breed in wool recording was Merino, followed by Dohne Merino and local breeds. The traits involved were fleece weight, clean fleece weight or yield, fibre diameter, fibre diameter variation, staple length, staple strength, homogeneity of fleece, fibre density, fibre curvature, colour, visual appreciation, and additional traits. The most frequently recorded phenotypes were fibre diameter, staple length, fleece weight, fibre diameter variation, visual appreciation, and colour. The survey results provided useful insight into wool recording and will form the basis for guidelines development.

Keywords: sheep, wool recording, on-line survey, traits, genetic evaluation

Introduction

The ICAR Sheep, Goats and Camelids Working Group (abbreviated SGC WG) has developed the guidelines and standards for performance recording in dairy sheep and goats and the guidelines for performance recording of growth, meat, reproduction and maternal ability in sheep and goats. To date, wool traits have not been included in the ICAR guidelines, although wool production is an important sector of genetic improvement and performance recording internationally. Contrary to countries that keep mainly dairy or meat sheep and goats (such as most European countries), the cost of sheep shearing is higher than the value of the wool, which means a financial loss for the breeder. Due to the increasing demand for recommendations on wool, ICAR has set itself the goal of expanding its guidelines to include the traits related to wool. For that purpose, an Expert Advisory Group was established to define the list of wool traits of interest, the methods of their recording, collecting, measuring, assessing, and to propose the method and statistical models for genetic evaluation. To achieve these objectives, an on-line survey on wool recording has been created. Relevant institutions were invited to complete the on-survey during the years 2020 and 2021 in order to collect the relevant information worldwide. The aim of this paper was to provide an overview of the feedback from the on-survey that will form the basis for recommendations on wool recording and genetic improvement of wool traits.

Material and methods

Breeding organisations and other relevant institutions responsible for sheep recording were invited to fill out an on-line survey about sheep wool. The survey includes basic information about the breeding organisations (name, country, and contact details for the organisation's representative) and specific questions regarding the wool population size (total sheep population size, wool sheep population size and number of farms with wool performance recording), information regarding wool breeds – the main breed and additional breeds if they are included in wool recording (total population size, population in wool performance recording, and number of farms with wool performance recording), and relevant information about recorded wool traits. The following list of wool traits has been proposed for recording: fleece weight, clean fleece weight or yield, fibre diameter, fibre diameter variation, staple length, staple strength, homogeneity of fleece, fibre density, fibre curvature, colour, other additional traits (such as different kind of visual appreciation), and traits requiring a sample of wool. For the latter, the protocol of sampling were requested. The following information's for traits have been collected from the survey: recording of the trait (yes/no), who collects the data, method of collecting, device for measuring, unit of trait expression, recording age of the animals, minimal/optimal wool growth period, assessment (subjective or measured), information's about genetic evaluation of the trait (genetic parameters, evaluation method and statistical models, inclusion of the trait in the economic index). Additional relevant data on management of animals (e.g., shearing) were also collected in the survey.

The SAS statistical package (SAS Inst. Inc., 2009) was used to analyse the data and generate descriptive statistics.

Results and discussion

In total, 17 relevant institutions on wool recording from 14 countries responded to the on-line survey (Table 1). Most of institutions (10 of them) answered the survey completely, while seven institutions completed partial survey.

As expected, the largest wool population is coming from Australia, followed by New Zealand and South Africa (Table 2). Only a small proportion of these populations (between 0.09 and 4%) are under wool performance recording. However, these countries have respectable number of farms with performance recording for wool. On the other hand, the entire wool population of Portugal is involved in the performance recording with a certain number of farms.

Table 1. Survey status by country (and breeding organisation).

Country	**Status of survey**	
	Fully completed	**Partially completed**
Australia		1
Austria		1
Bulgaria		1
Croatia		1
Czech Republic	1	
Finland	1	
Iceland	1	
Latvia	1	
New Zealand	1	1
Portugal	1	1
Slovenia	1	
South Africa	2	
Sweden		1
Uruguay	1	
All	10	7

Table 2. Total size of the wool sheep population, population size, number (N) of wool performance-recorded ewes (WPR), proportion (%) of population in WPR and number of farms (no.) in WPR.

Country	**Total size**	**N in WPR[1]**	**% in WPR**	**No of farms in WPR**
Australia	70,000,000	2,900,000	4	222
Bulgaria	10	2	20	1
Croatia	42,000	200	0.47	1
Finland	14,170	1,000	7	10
Latvia	28,224	5,448	19	49
New Zealand	27,600,000	20,000	0.7	35
Portugal	26,232	26,232	100	90
Slovenia	5,452	150	2.7	2
South Africa	24,000,000	626,000	3	589
South Africa	15,000,000	13,550	0.09	65
Uruguay	6,723,548	25,500	0.4	81

[1] WPR = wool performance recording

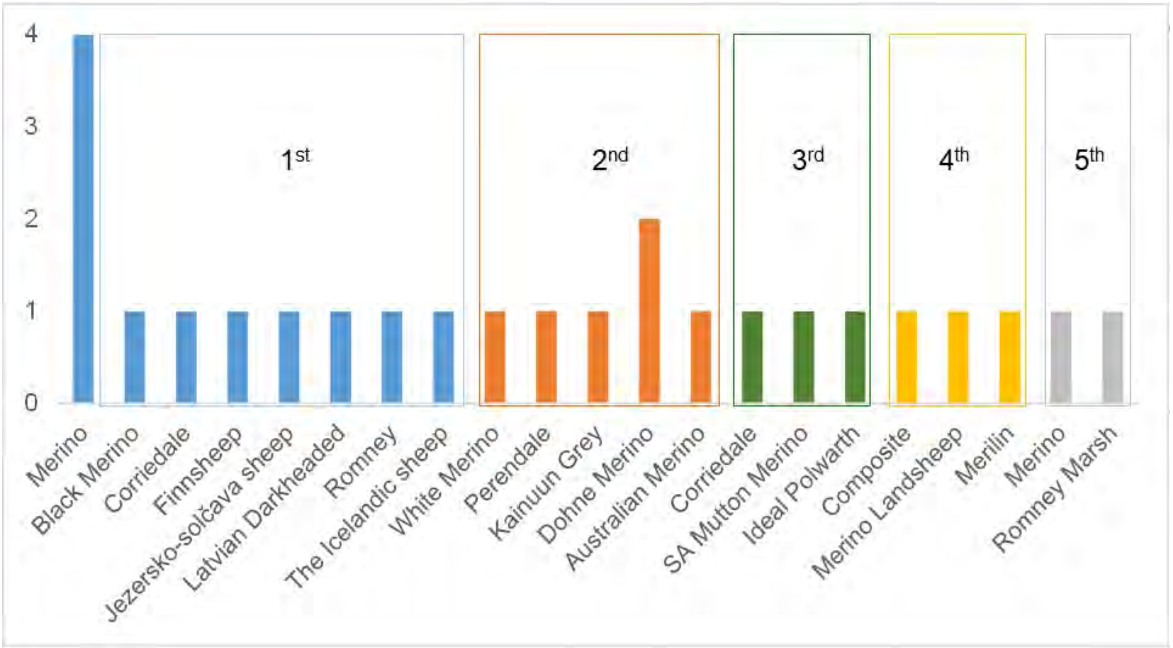


Figure 1. Frequency distribution of common breeds under wool performance recording.

The next part of the survey was related to the breeds having wool performance recording (Figure 1) and 11 of the 17 breeding organisations responded. As expected, the most common main breed was Merino (36%) followed by Dohne Merino and local breeds. There was also a possibility to include additional breeds in the survey (from second to fifth breed). Besides specialised wool breeds, some of them belong to the group of mountain and lowland breeds. Both, males and females were included in the wool sample collection.

The most common traits for wool performance recording are reported to be fibre diameter, staple length, and fleece weight (Figure 2). Beside these, fibre diameter variation, visual appreciation, and colour have been traits frequently used in wool recording. Less frequent traits were staple strength, fibre density and homogeneity. Beside these traits, organisations have the opportunity to provide information about additional traits if they were recorded but not included in the given list (five responses). On-line survey also offered the possibility to provide information about traits that require a wool sample (seven responses).

Respondents were asked about the person who collecting the data, the method and the device used for measuring, and the units of trait expression. The results regarding data collection are summarised in Figure 3. The farmer was most common person involved in data collection, followed by technicians and laboratory staff. Experts were involved in the collecting specific traits such as visual appreciation. For traits requiring practical knowledge (e.g., fibre curvature, staple strength, fibre density, and homogeneity), data collection was performed by technicians.

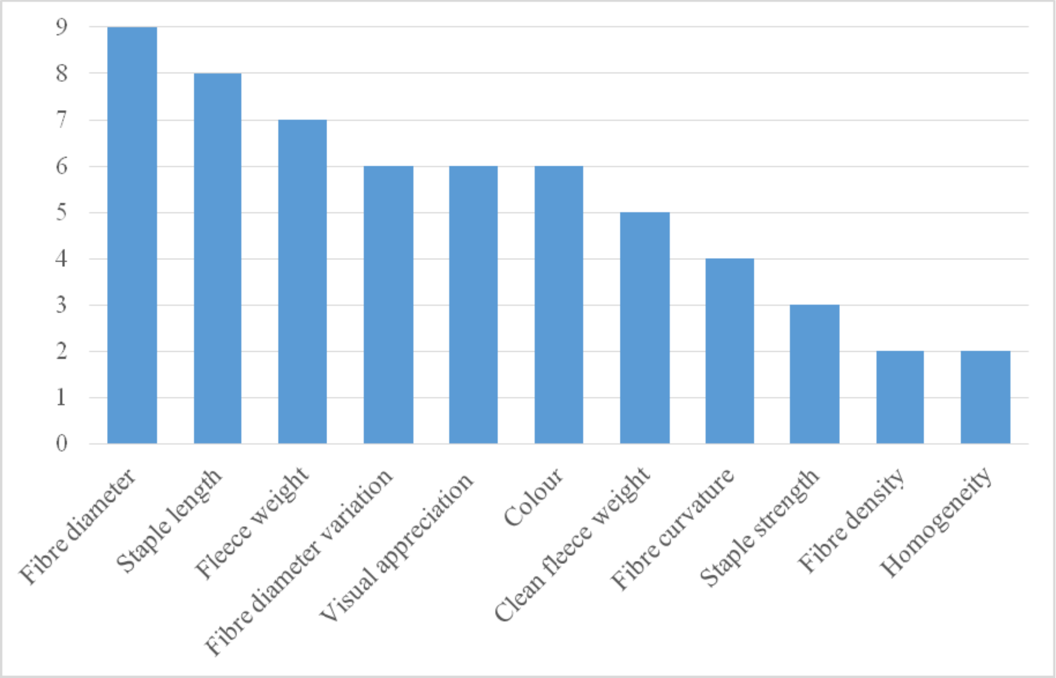


Figure 2. Frequency distribution of traits involved in the wool performance recording.

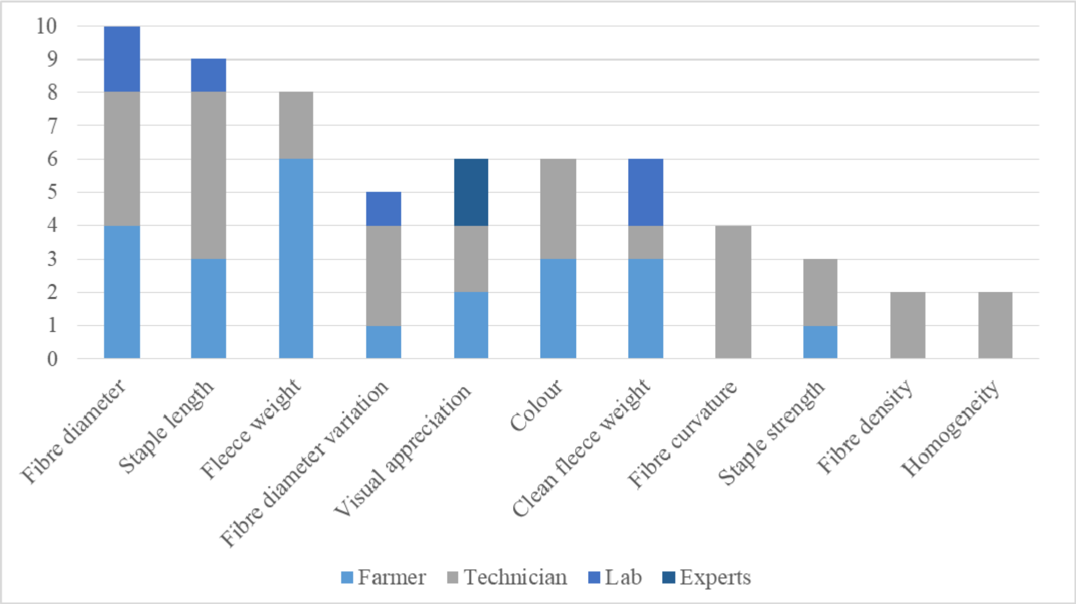


Figure 3. Persons responsible for data collecting by trait.

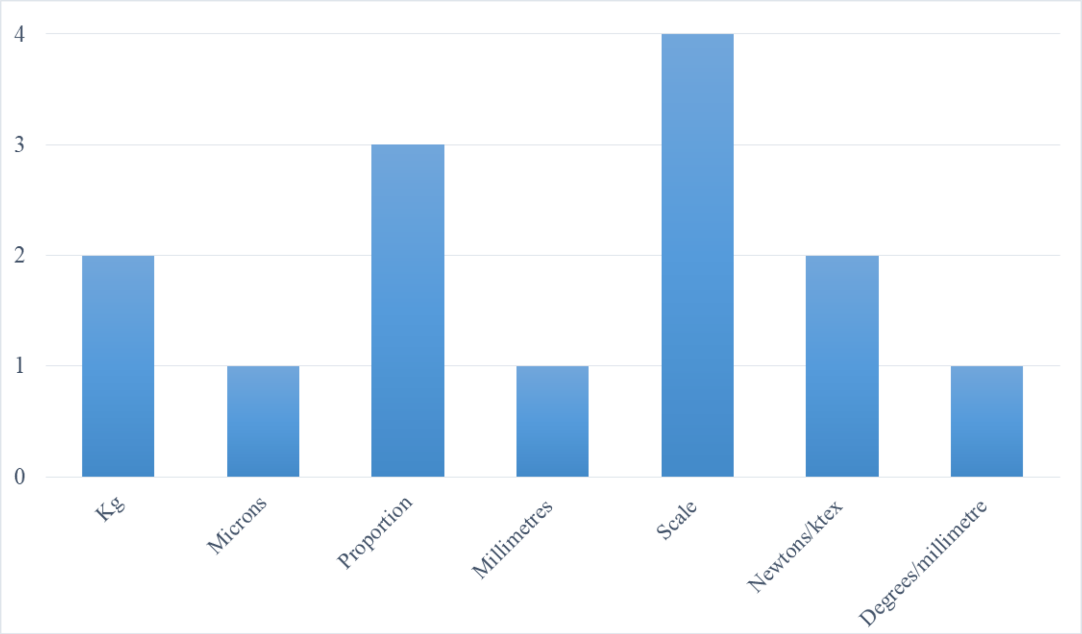


Figure 4. Unit of trait expression.

Methods used for recording wool are trait related. According to the respondents, some traits are weighed (fleece weight, clean fleece weight), others are either visually recorded (colour, fibre density, homogeneity), scored by scale (visual appreciation), tested with different methods (atlas or pull test have been used for staple strength recording), or measured with special machines, instruments and devices (fibre diameter, fibre diameter variation, staple length, and fibre curvature). In the case when traits are weighed or measured, the equipment used for this purpose was also specified. The units for trait expression (Figure 4) were either kg (fleece weight, clean fleece weight), microns (fibre diameter), proportions (clean fleece weight, fibre diameter variation), millimetres (staple length), Newtons/ktex (staple strength), scale (homogeneity, fibre density, colour, visual appreciation), and degrees/millimetre (fibre curvature).

Results regarding recording age and minimal/optimal growth period showed wide variation depending on trait, breed, and country. Recording age ranged from three (fibre curvature) to 28 months (fleece weight and clean fleece weight), while growth period ranged from three (fleece weight, fibre diameter) to 18 months (visual appreciation).

The last part of the survey included information regarding genetic evaluation. In most of the responding organisations, breeding values for wool traits have been estimated (Figure 5). The exceptions are fibre density and homogeneity. BLUP animal model as computation method was considered as being the standard for the genetic evaluations. In addition, the BLUP method has been extended to GBLUP by including molecular data for genomic evaluation of wool traits in Australia. Proposed traits have been mostly included in the economic index. The survey also included additional information on traits such as average phenotype, annual genetic trend, and genetic parameters (genetic variance and heritabilities) to provide an overview of how heritable these traits are.

]

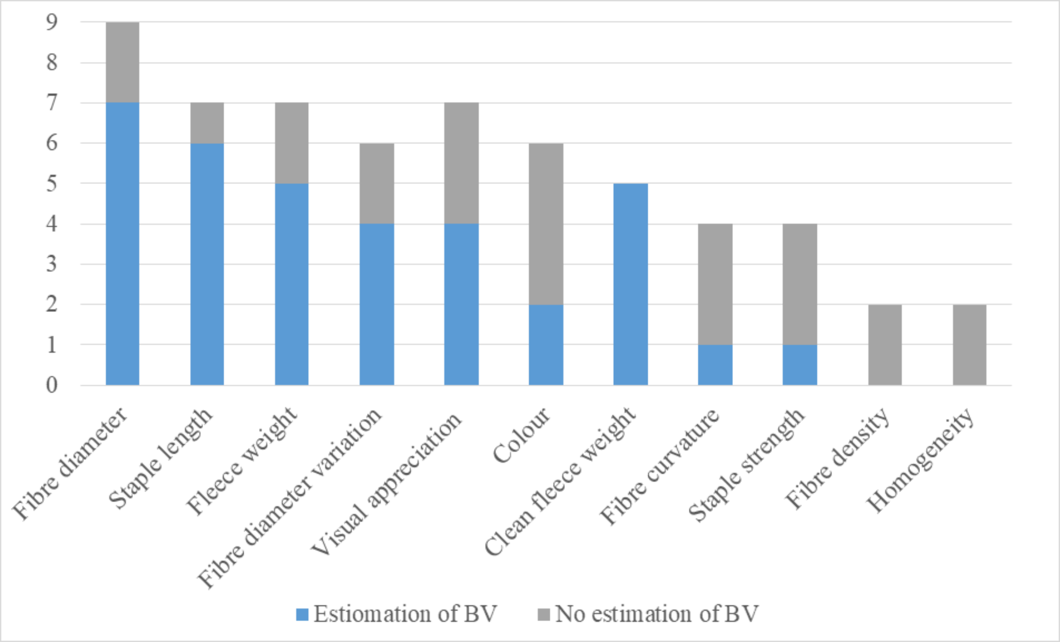


Figure 5. Estimation of breeding value by trait.

Additional relevant data on management of animals (e.g., shearing) were also collected in the survey. Responses indicate that animals are shared once or twice a year, usually in the spring and autumn. The final element of the survey, an open textbox option allowed for qualitative feedback that would be useful for completeness of the survey. Most respondents chose to tell more about their breeds and related issues.

Conclusion

The results from the on-line survey offer useful insight in wool recording and provide some general considerations, common traits that could be collected and recommendations for wool performance recording. They also provide information about genetic evaluation as a tool for genetic improvement of wool production traits. Furthermore, the survey results will serve as a basis for developing new ICAR guidelines for recording wool traits in sheep.

References

SAS Inst. Inc. 2009. SAS/STAT® 9.2 User's Guide, Cary, NC: SAS Institute Inc

The implementation of the standard lactation length and standard milking length in milk recording of dairy sheep and goats breeds in Slovenia

M. Bizjak and M. Simcic

University of Ljubljana, Biotechnical Faculty, Department of Animal Science,
Jamnikarjeva 101, 1000 Ljubljana, Slovenia

Corresponding author: mojca.simcic@bf.uni-lj.si

Abstract

The aim of the study was to introduce the standard lactation length and standard milking length in milk recording of Slovenian dairy sheep and goat breeds in breeding programs to achieve better comparability among animals within the same breed in milk production traits (total milk yield, total milked milk as well as protein, fat and lactose contents). In Slovenia, there are two local dairy sheep breeds, Bovec sheep and Istrian Pramenka and an Improved Bovec sheep. There are three dairy goat breeds, a local Dreznica goat and two locally adapted breeds, Slovenian Alpine goat and Slovenian Saanen goat. According to the recommendations in the ICAR guidelines, every approved organisation should define the standard lactation length (and/or standard milking length), which is close to the average lactation length (milking length) of the considered breed according to the production system. At the beginning, an average age at weaning of offspring (beginning of milking) and average milking length (end of milking) were calculated. Based on these calculations the standard lactation length and standard milking length were defined. Due to two different production systems used for Slovenian Alpine and Slovenian Saanen goats, animals of both breeds were divided in two subpopulations. In the first one, where kids are early weaned (until the 5th day of age) it was determined that standard lactation length lasts from the day of kidding to 240th day after kidding in both breeds. In the second subpopulation, where kids are weaned later than the 5th day of age (late weaning) we defined a standard milking length from 40th to 240th day after kidding. In the case of Dreznica goat and in all Slovenian dairy sheep breeds early weaning practically does not exist. Consequently, only a standard milking length was defined. The standard milking length in Dreznica goat, Bovec sheep and Improved Bovec sheep was in the period from 40th to 210th day after giving birth, while in the case of Istrian Pramenka from 60th to 210th day after lambing. In the future, we will continue with monitoring of total milk yield, total milked milk as well as protein, fat and lactose contents produced in the standard lactation length and in the standard milking length according to the breeding program of each breed and adjust if necessary.

Keywords: Standard lactation length, standard milking length, dairy sheep, dairy goats, Slovenia.

Introduction

The lactation length is one of the main factors, which affect total milk yield and milk composition in dairy cattle, sheep and goats, and it is often used as one of the criteria for comparison among animals within the same flock, population or breed. However, milk production traits are easier to compare among animals when a milking length is of the same length. For this reason, standard lactation length is used in dairy cattle breeding for many years and it is accepted worldwide. It lasts from calving to the 305th day of lactation in all cattle breeds, while there is no such uniform standard lactation length in dairy sheep and dairy goats. However, in several breeding programs around the world standard lactation length is used to compare animals within different sheep and goat breeds. In the breeding programs for German Alpine goat (Zuchtprogramm Bunte Deutsche Edelziege, 2021) and German Saanen goat (Zuchtprogramm Weise Deutsche Edelziege, 2021) a standard lactation length lasts 240 days, while in the breeding program for the East Friesian sheep in Germany (Zuchtprogramm Ostfriesisches Milchschaf, 2021) the dairy traits are recorded in a period of 150 days. In the guidelines for the implementation of milk recording in sheep and goats in Austria, a standard lactation length is determined as a period from giving birth to the 240th day of lactation (LKV Austria, 2022). One of the most important criteria in the implementation of standard lactation length is the average lactation length of each breed. Some researchers also considered some other criteria such as genetic parameters and genetic correlations between milk production traits in the standard lactation and milk production traits in the whole lactation. Based on described criteria, Basdagianni *et al*. (2018) determined that standard lactation length in the Chio diary sheep lasts 220 days or 190 days in ewes in the first lactation, respectively. The implementation of standard lactation length is reasonable in the intensive production systems where kids and lambs are weaned in the first few days after birth. Consequently, the milking length is practically identical to the whole lactation length. On the other hand, in less intensive systems, lambs and kids are weaned later, and the milking length is shorter than the whole lactation length. In the extensive production systems, the implementation of the standard milking length (a period between the weaning of offspring and the end of lactation) is more appropriate than the implementation of the standard lactation length.

The aim of the study was to introduce the standard lactation length or standard milking length, respectively, in the milk recording of Slovenian dairy sheep and goat breeds in the breeding programs to achieve better comparability among animals within the same breed in the milk production traits (total milk yield, total milked milk as well as protein, fat and lactose contents).

Material and methods

Records were provided by the Slovenian breeding programs for dairy sheep and dairy goats collected from the year 2005 to 2019. Data of animal birth date, breed, flock, lambing/kidding date, parity, weaning date, date of the end of lactation and records of milk recording were obtained from the Central Database for Small Ruminants in Slovenia. Only records with the ewes'/does' lactation length between 0 and 300 days and lambs'/kids' suckling period length between 0 and 100 days were included in the analysis. Likewise, ewes and does, which had less than three milk recordings within the same lactation, were excluded from the analysis. Sheep flocks with less than 50 records as well as goat flocks with less than 30 records were excluded as well. After the records control, the number of included records was 27,234 for dairy sheep and 12,289 for dairy goats.

Based on these records, the average age at weaning of offspring (beginning of the milking) and the average milking length (end of milking) for each breed were calculated according to the ICAR guidelines which recommends that each approved organisation should define the standard lactation length (and/or standard milking length) which is close to the average lactation length (average milking length) of the considered breed according to the production system (ICAR, 2018). Due to two different production systems of Slovenian Alpine goat and Slovenian Saanen goat in Slovenia, animals of both breeds were divided into two subpopulations. In the approximately 25% of flocks of these two breeds kids are early-weaned (till the 5th day after birth), immediately after the end of the colostral phase of suckling. Especially in the case of Slovenian Saanen goat, the percentage of flocks where early weaning occurs has increased in the last few years. In the first subpopulation, standard lactation lasts from the day of kidding to the 240th day after kidding in both breeds. In the second subpopulation, where kids are weaned later than the 5th day of age (late weaning), a standard milking length from 40th to 240th day after kidding were determined. In the case of Dreznica goat and all dairy sheep breeds in Slovenia, early weaning is practically not in use. Consequently, for these breeds, only a standard milking length was defined. The standard milking length in Dreznica goat as well as in Bovec sheep and Improved Bovec sheep was agreed in the period from 40th to 210th day after giving birth, while in the case of Istrian Pramenka it lasts from 60th to 210th day after lambing.

The total milk yield in the standard lactation length and the total milked milk in the standard milking length are calculated from the records obtained in milk recordings where daily milk yield is measured with milk meters and a milk sample is taken for the determination of protein, fat and lactose contents in the milk of each ewe or doe in a flock. The equation 1 is used for the total milk yield and the total milked milk estimation based on the daily milk yields from milk recordings and intervals between two successive milk recordings:

$$TMY\ (TMM) = (I_0 * M_1 + I_1 * (M_1 + M_2) / 2 + ... I_n * M_n) / 1000$$

where:

TMY = total milk yield in the standard lactation length (kg),

TMM = total milked milk in the standard milking length (kg),

I_0 = the interval between the beginning of standard lactation/standard milking and the first milk recording (days),

$M_1, M_2, \ldots M_n$ = daily milk yield (g),

$I_1, I_2 \ldots$ = the interval between two successive milk recordings (days),

I_n = the interval between the last milk recording and the end of the standard lactation length/standard milking length (days).

The protein, fat and lactose yields in the standard lactation length or in the standard milking length are estimated from the daily milk yields multiplied by the protein, fat and lactose contents determined in the milk sample taken at each milk recording divided by 100.

The total milk yield in the standard lactation length and the total milked milk in the standard milking length are estimated just for ewes and does which comply with the following conditions:

- an animal had at least three milk recordings per lactation; two of them were within the standard lactation length or within standard milking length;
- first milk recording must take place within 52 days after the weaning of offspring;
- the average interval between two successive milk recordings is within a range from 28 to 34 days;

- there is a tolerance of one missed milk recording per lactation as well.

Results

Descriptive statistic for the total milked milk and milk composition of dairy sheep breeds in the standard milking length from the year 2015 to 2021 is presented in Table 1. Improved Bovec sheep had the highest average total milked milk and the lowest average protein and fat contents (191.9 kg, 5.2% and 5.7%), respectively. On the other hand, Istrian Pramenka had the lowest average total milked milk and the highest average protein and fat contents (99 kg, 5.9% and 7.2%), respectively.

In the Table 2 is descriptive statistics for the total milked milk and milk composition of dairy goat breeds in the standard milking length in the years 2015-2021. On average, the total milked milk was the highest in Slovenian Saanen goat (376.9 kg) and the lowest in Dreznica goat (267.7 kg). On the contrary, the average protein, fat and lactose contents were the highest in the milk of Dreznica goat (3.4%, 4.1% and 4.5%), respectively and the lowest in the milk of Slovenian Saanen goat (2.9%, 3.0% and 4.2%), respectively.

Descriptive statistic for the total milk yields and milk composition of dairy goat breeds in the standard lactation length in the years 2015-2021 is presented in Table 3. Slovenian Saanen goat had higher average total milk yield (517.0 kg) and lower average protein and lactose contents (2.9% and 4.4%) in comparison with Slovenian Alpine goat (511.6 kg, 3.2% and 4.5%), respectively, while the average fat content was the same in both breeds (3.2%).

Discussion

The goal of this study was to define and to introduce the standard lactation length and the standard milking length appropriate for each breed of dairy sheep and dairy goats in Slovenia. In the estimation of the standard lactation length and standard milking length the production system of ewes and does was also considered. Although the standard lactation length or standard milking length in sheep and goats is primarily used as criteria for comparison among animals within the breed, in some cases animals of different breeds could be compared as well. The Bovec sheep is possible to compare to the Improved Bovec sheep due to the same standard milking length. However, the Improved Bovec sheep is a result of improving Bovec ewes with East Friesian rams. A higher total milked milk in the Improved Bovec sheep on average (191.9 kg) was expected compared to Bovec sheep (137.7 kg), which means that improving had a positive effect on the milk yield. In the same time, the Improved Bovec sheep had lower average dry matter content (sum of protein, fat and lactose contents; 15.5%) than Bovec sheep (16.3%) which is a consequence of higher milk yield in Improved Bovec sheep. Istrian Pramenka had the lowest average total milk yield (99.9 kg) and the highest dry matter content (17.6%) in the standard milking length than other two breeds. However, the total milked milk of all three breeds could not be compared because the standard milking length in Istrian Pramenka is 20 days shorter in comparison with Bovec sheep and Improved Bovec sheep.

The subpopulations of Slovenian Saanen goat and Slovenian Alpine goat where kids are late-weaned are compared to each other because of the same standard milking length. Slovenian Saanen goat had higher total milked milk on average (376.9 kg) than Slovenian Alpine goat (323.9 kg). Likewise, Slovenian Saanen goat had lower average dry matter content (10.1%) in comparison with Slovenian Alpine goat (10.6%). Dreznica goat had 267.1 kg of milk and it had the highest average dry matter content (12.0%) in the standard milking length than Slovenian Saanen goat and Slovenian Alpine goat

Table 1. Descriptive statistics for the total milked milk and milk composition of dairy sheep breeds in the standard milking length from 2015 to 2021.

Breed/ Standard milking length	Variable	N^1	Mean	SD^2	Min	Max
Bovec sheep 40th to 210th day	Total milked milk (kg)	8,750	137.7	53.7	12.4	450.7
	Protein content (%)	8,672	5.3	0.4	3.3	10.1
	Fat content (%)	8,672	6.4	0.8	3.1	13.9
	Lactose content (%)	8,672	4.6	0.3	2.3	7.9
Improved Bovec sheep 40th to 210th day	Total milked milk (kg)	3,128	191.9	75.1	21.0	514.7
	Protein content (%)	3,127	5.2	0.4	3.9	7.8
	Fat content (%)	3,127	5.7	0.8	2.8	9.7
	Lactose content (%)	3,126	4.6	0.3	2.3	6.9
Istrian Pramenka 60th to 210th day	Total milked milk (kg)	1,972	99.9	31.5	22.3	350.2
	Protein content (%)	1,922	5.9	0.5	4.5	8.3
	Fat content (%)	1,922	7.2	0.8	3.4	10.8
	Lactose content (%)	1,922	4.5	0.3	1.9	6.5

[1] Number of records.
[2] SD – Standard Deviation.

Table 2. Descriptive statistics for the total milked milk and milk composition of dairy goat breeds in the standard milking length from 2015 to 2021.

Breed/ Standard milking length	Variable	N^1	Mean	SD^2	Min	Max
Slovenian Saanen goat 40th to 240th day	Total milked milk (kg)	1,107	376.9	127.3	51.9	954.4
	Protein content (%)	1,107	2.9	0.3	2.2	4.1
	Fat content (%)	1,106	3.0	0.6	1.5	5.9
	Lactose content (%)	1,107	4.2	0.2	3.6	4.8
Slovenian Alpine goat 40th to 240th day	Total milked milk (kg)	2,602	323.9	153.0	12.7	1102.2
	Protein content (%)	2,595	3.1	0.3	2.2	5.3
	Fat content (%)	2,595	3.2	0.6	1.6	7.4
	Lactose content (%)	2,595	4.3	0.2	3.2	6.6
Dreznica goat 40th to 210th day	Total milked milk (kg)	1,972	267.1	129.5	41.6	833.2
	Protein content (%)	1,922	3.4	0.3	2.7	5.6
	Fat content (%)	1,922	4.1	0.7	2.1	7.5
	Lactose content (%)	1,922	4.5	0.2	3.6	7.1

[1] Number of records.
[2] SD – Standard Deviation.

Table 3. Descriptive statistics for the total milk yields and milk composition of dairy goat breeds in the standard lactation length in the years 2015-2021.

Breed/ Standard lactation length	Variable	N^1	Mean	SD^2	Min	Max
Slovenian Saanen goat 1st to 240th day	Total milk yield (kg)	1,040	517.0	196.4	66.0	1,385.3
	Protein content (%)	1,039	2.9	0.2	2.3	3.9
	Fat content (%)	1,039	3.2	0.5	1.8	6.0
	Lactose content (%)	1,039	4.4	0.2	3.7	5.4
Slovenian Alpine goat 1st to 240th day	Total milk yield (kg)	694	511.6	177.4	76.8	1,219.8
	Protein content (%)	694	3.2	0.3	2.4	4.4
	Fat content (%)	694	3.2	0.4	1.5	4.5
	Lactose content (%)	694	4.5	0.2	4.0	4.9

[1] Number of records.
[2] SD – Standard Deviation.

(subpopulations with late weaning of kids). However, a direct comparison of the total milked milk of all three goat breeds is not reasonable, because the standard milking length in Dreznica goat lasts 30 days less compared to other two goat breeds.

The total milk yield of Slovenian Saanen goat and Slovenian Alpine goat (subpopulations with early-weaned kids) could be compared due to the same standard lactation length. Slovenian Saanen goat had a little higher average total milk yield (517.0 kg) and a lower average dry matter content (10.5%) than Slovenian Alpine goat (511.6 kg and 10.9%). The standard lactation length in both breeds lasts from kidding to the 240th day of lactation, which is the same as in breeding programs for German Saanen goat (Zuchtprogramm Weise Deutsche Edelziege, 2021) and German Alpine goat (Zuchtprogramm Bunte Deutsche Edelziege, 2021) in Germany. For this reason, a direct comparison of total milk yield in considered Slovenian and German breeds is possible. In Slovenian Saanen goat, the average total milk yield in the standard lactation length (517.0 kg) was lower in comparison with German Saanen goat (750 – 1,000 kg). On the other hand, the protein (2.9%) and fat (3.2%) contents in the milk of Slovenian Saanen goat were similar to the milk of German Saanen goat (2.8 - 3.0% protein content and 3.2 – 3.5% fat content). In the Slovenian Alpine goat, the average total milk yield in the standard lactation length (511.6 kg) was lower in comparison with the German Alpine goat (850 – 1,200 kg). Fat content (3.2%) in Slovenian Alpine goat was in accordance with the fat content in the milk of German Saanen goat (3.2 – 3.5%), while the protein content in the milk of Slovenian Alpine goat (3.2%) is even above the protein content in the milk of German Saanen goat (2.8 – 3.0%).

Conclusion

In the year 2022, we have introduced the standard lactation length and standard milking length in milk recording of Slovenian dairy sheep and goat breeds in breeding programs to achieve better comparability in the milk production traits among animals within the same breed. In the future, we will continue with the monitoring of total milk yield and total milked milk as well as protein, fat and lactose contents in the standard lactation length and standard milking length according to each breed and adjust if necessary.

References

Basdagianni, Z., E. Sinapis, G. Banos, 2019. Evaluation of reference lactation length in Chios dairy sheep. Animal 13(1): 1-7.

ICAR, 2018. Guidelines for Performance Recording in Dairy Sheep and Dairy Goats. ICAR, Utrecht, The Netherlands, 37 pp. https://www.icar.org/Guidelines/16-Dairy-Sheep-and-Goats.pdf

LKV Austria, 2022. Richtlinien der Milchkontrolle Schafe und Ziegen. LKV Austria, Wien, Austria https://lkv.at/at/leistungspruefung/themen/milchleistung/Schafe-und-Ziegen/Richtlinien-Milchleistungspruefung-Schafe-Ziegen.php

Zuchtprogramm Bunte Deutsche Edelziege, 2021. Ziegenzuchtverband, Stuttgart, Germany, 5pp. Microsoft Word - ZP BDE BW 18_01_2022.docx (verwaltungsportal.de)

Zuchtprogramm Ostfriesisches Milchschaf, 2021. Landesschafzuchtverband, Stuttgart, Germany, 5pp. Zuchtprogramm Ostfriesisches Milchschaf – Landesschafzuchtverband Baden-Württemberg e.V. (schaf-bw.de)

Zuchtprogramm Weise Deutsche Edelziege, 2021. Ziegenzuchtverband, Stuttgart, Germany, 5pp. Microsoft Word - ZP WDE BW 18_01_2022.docx (verwaltungsportal.de)

Genetic improvement: A fundamental pillar for the canadian sheep industry

F. Fortin[1], M. Trottier-Lavoie[1], A. St-Pierre[1] and F.Schenkel[2]

[1]Centre d'expertise en production ovine du Québec, 198, avenue Industrielle, La Pocatière, QC, Canada
Corresponding Author: frederic.fortin@cepoq.com
[2]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada

Abstract

The Canadian sheep industry is a relatively small industry and primarily focused on meat production. GenOvis is the national genetic evaluation system with close to 200 participants in 2022 and more than 1 million lamb records. The main purebred breeds under genetic selection are the maternal Dorset and Polypay breeds, the prolific Romanov and Rideau Arcott breeds and terminal Suffolk and Hampshire breeds. In the past years, genetic gains have been achieved for multiple traits using selection indexes. From 2012 to 2021, the largest realized total genetic gains in growth and carcass traits were made by Suffolk breed for 50-day weight (+1.18kg), gain from 50 to 100 days (+2.21kg) and the loin depth (+1.56 mm). In addition, for maternal traits, the largest genetic gains have been observed in Polypay breed for maternal 50-day weight (+0.84 kg) and in Romanov breed for total weaning weight at later parities (+2.55 kg) and total number born at later parities (+0.11 lambs). During the same period, the change in inbreeding levels varied from negative in Polypay (-0.2%) to the largest positive increase in Dorset (+1.1%). There are many opportunities to accelerate these genetic gains. The increase of data collection through multiplier and commercial producers, new phenotyping technologies, and integration of genomic information are among the most promising developments.

Keywords: Genetic gain, sheep breeding program, selection traits, inbreeding

Introduction

The Canadian sheep industry is a relatively small industry compared to chickens, pigs or cattle. As for 2019, 827,800 sheep were produced in Canada compared to 171 398 000 chickens, 14 399 300 pigs and 11 500 000 cattle (Agriculture Canada, 2021). The Canadian sheep industry is primarily meat focused, but milk and wool are also produced. Overall, lamb meat represented 4% of the total retail sales value market of Canadian meat sector in 2019 (Agriculture Canada, 2021). Moreover, the Canadian lamb industry compound annual growth rate (CAGR) has been increasing of 0.1% from 2015 to 2019 (Agriculture Canada, 2021). This upward trend is continuing as the CAGR from 2019 to 2023 is predicted to increase of 3% (Agriculture Canada, 2021). Canadian sheep industry is becoming more and more important and constant efforts must be maintained to specialize it. To do so, tools are made available to sheep producers like for instance the GenOvis program. The GenOvis program has been created as a mean to give an on-farm sheep genetic program for all Canadian users, namely purebred sheep breeders, commercial producers, and dairy farmers. This program+ a partnership between Centre d'expertise en production ovine du Québec (CEPOQ), Ontario Sheep

Farmers (OSF), Canadian Sheep Breeders Association (CSBA), and the Centre for Genetic Improvement of Livestock (CGIL) and counts 199 members divided into 166 breeders, 15 commercial producers, and 18 dairy producers. The technical support is done by CEPOQ whereas the web application and genetic evaluation routines are developed and run by CGIL. Two modules are offered which are lamb (meat), since 2000, and dairy module, since 2014. These modules help sheep farmers to produce high quality lambs for consumers, but also gave access to genetic values that enable them to improve their ewes milk production and quality. Overall, the program allows an effective evaluation of sheep genetic value based on important economic traits as expressed by relatives and other animals within breed or crossbreed.

Material and methods

This paper describes and presents results only about the GenOvis lamb (meat) module. A total of 15 traits are recorded in the database from which three groups emerge, namely growth, carcass, and reproduction (Table 1).

Birthweight is used to calculate adjusted 50 days weights and is considered valid only when it is between 0.5 kg to 9.9 kg (Schaeffer and Szkotnicki, 2015). Thereby birthweight outside this range remains in the database but is set to missing for genetic

Table 1. Description of the recorded traits in GenOvis evaluations.

Group	Traits
Growth	Lamb survival (direct) (%) Birth weight (direct) (kg) 50-day weight (direct) (kg) Gain 50 – 100 days (kg)
Carcass	Fat depth – ultrasound (mm) Loin depth – ultrasound (mm)
Reproduction	Lamb survival (maternal) (%) Birth weight (maternal) (kg) 50-day weight (maternal) (kg) Age 1st lambing (days) Lambing interval (days) # Born (1st and later lambings) Total weaning weight (1st and later lambings)

Table 2. Average birth weight used by breed group when birth weight is missing.

Breed Group	Average birth weight (kg)
1	3.67
2	4.53
3	4.86
4	4.11
5	4.98
6	4.72
7	4.83
8	2.83
9	4.17
10	3.88
11	4.04
12	4.57
13	3.77
14	4.48
15	3.75

evaluation purposes (Schaeffer and Szkotnicki, 2015). When not reported, an average birthweight per breed is used (Table 2) (Schaeffer and Szkotnicki, 2015).

50-day weight is the weight taken around 50 days of age, which, at a practical level, means that the actual weight should be taken between 28 to 72 days of age (Schaeffer and Szkotnicki, 2015). If the age at weighing is outside this range and/or the taken weight is not between 4.0 kg and 51.0 kg, the weight is declared missing (Schaeffer and Szkotnicki, 2015). As growth from birth to 50 days of age is assumed to be linear, weight is adjusted to 50 days of age by extrapolation (either up or down) (Schaeffer and Szkotnicki, 2015).

Lamb survival is defined with 5 categories, where 1 is an animal that died shortly after birth and 5 an animal that has made it to weaning age (Schaeffer and Szkotnicki, 2015). Categories 2, 3 and 4 are different lengths of survival, but with a death before weaning (Schaeffer and Szkotnicki, 2015).

Gain from 50 to 100 days is the difference between the weight at 100 days of age and 50 days of age (Schaeffer and Szkotnicki, 2015). The 100-day weight is provided by producers, which should be taken between 73 to 135 days of age and should be between 10.0 kg to 80.0 kg (Schaeffer and Szkotnicki, 2015). As growth from 50 days to 100 days of age is assumed to be linear, 100d weight is adjusted to 100 days of age by extrapolation (either up or down). This is the gain 50-100d that is used in breeding values calculation. The 100-day weight must be taken at least 28 days after the weaning weight otherwise it is set to missing. (Schaeffer and Szkotnicki, 2015).

Fat and loin depth are ultrasound traits, which means that they are taken by ultrasound measurements (Schaeffer and Szkotnicki, 2015). Ultrasound measurements are taken around 100 days of age (between 73 to 135 days of age) (Schaeffer and Szkotnicki, 2015). Lambs must be weighed at the same time. The ultrasound weights must be between 10.0 kg to 80.0 kg. For the measurement lambs must stand upright on all four legs and kept motionless in a cage, a balance or other installations (Fortier, 2022). Lambs back must be straight (Fortier, 2022). The measured zone is located between the 3rd and 4th lumbar vertebrae (Fortier, 2022). The measured zone needs to be shaved and cleaned to enhance contact between the skin and the ultrasound standoff and thus improve image quality (Fortier, 2022). Measurement is taken perpendicular to fat layers and at the muscle deepest part (Fortier, 2022). Two thicknesses of fat are measured, a first one above the deepest part of the muscle, and a second one at 1.5 cm from the first measurement, always perpendicular to the layer of fat (Fortier, 2022). A third measurement is taken to establish the muscle depth (Fortier, 2022). The skin should not be included in the lamb fat thickness measurement (Fortier, 2022). Fat measurements must be higher than 0.0 mm and below 14.9 mm to be included in breeding values. Loin depth must be between 10.0 mm to 44.0 mm. Loin and fat measurements are adjusted on lamb weight instead of lamb age. This adjustment is calculated at breeding value calculation. Records out of range are set as missing values.

The total weaning weight (TWW) is the sum of the adjusted 50-day weights of lambs that were weaned, which is calculated for first and later lambing (Schaeffer and Szkotnicki, 2015). This includes number of lambs weaned, but also how well the raise dam contributed to the lambs weaning weight. The genetic model account for the proportion of male and female progeny as there is a sex difference in weaning weights to consider. Bottle raised lambs were ignored. (Schaeffer and Szkotnicki, 2015).

The EBVs (Estimated Breeding Values) are calculated across breeds by multi-traits models on a weekly basis using Fortran programs specifically adapted for the sheep context. The EBVs are transformed and published in Expected Progeny Difference (EPD) for the use of the breeders. The genetic trends presented below are the average EBVs of lamb born during each year from the genetic evaluation run of the second week of May 2022. There are 2 genetic evaluation models, one for growth and carcass traits and the other for reproduction traits. Six traits are currently evaluated for growth and carcass (survival, birthweight, 50-days weight, gain from 50 to 100 days, fat and loin depth). All 6 traits are evaluated for direct (dir) and maternal (mat) genetic effects even if maternal effects are low for gain, loin and fat. The EBV dir (e.g.: Birth weight dir) refers to the animal own ability to perform while the EBV mat (e.g.: Birth weight mat) refers to the dam contribution on a lamb performance. The recording allows the distinction of three (3) different dams that have influenced their trait observations. The first is the genetic dam of the lamb. This is the female that has provided one half of the animal's DNA. The second dam carries the embryo and gives birth to the lamb(s). In most cases the `birth' dam is the same animal as the `genetic' dam. However, for producers that use embryo transfer (ET), then `birth' dam may be an unrelated female. The third dam raises the lamb from birth to weaning age. The `raise' dam may be the same as the genetic dam or may be the same as the `birth' dam, or may be a foster dam. The `raise' dam may actually be a bottle for bottlefed lambs. Biologically, the `birth' dam has an influence on survival and birthweight. The `raise' dam has an influence on 50-d weights, gain, loin, and fat thickness. The genetic dam has an influence on all six traits through the relationship matrix. Similarly, litter effects are associated with either the `birth' dam or the `raise' dam. Age of dam effects are also assigned according to `birth' dams or `raise' dams (Schaeffer and Szkotnicki, 2015).

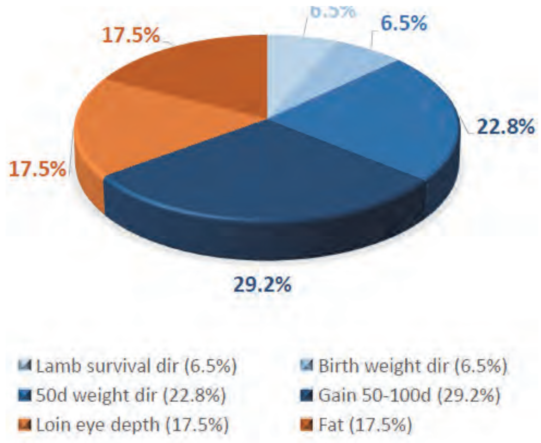
For the reproduction genetic evaluation model, a six-trait system was developed with three traits for parity1 and three traits for later parities. Age at first lambing, number born at first lambing, and TWW at first lambing were three traits for first parity animals, and interval between lambings, number born, and TWW were for second and later parities. First parity and later parity traits have a genetic correlation of only 0.7, so that they can be considered as different traits (Schaeffer and Szkotnicki, 2015).

Selection indexes

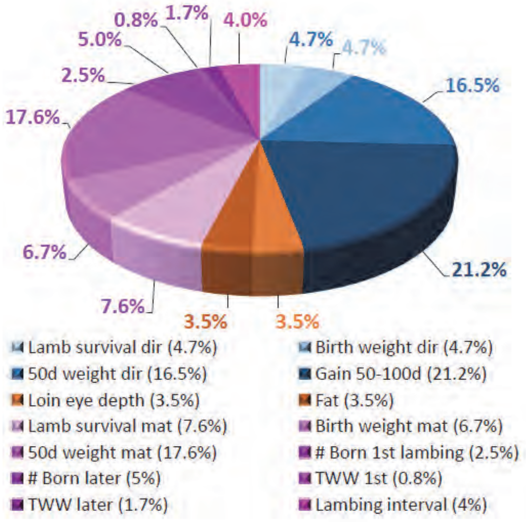
EBVs are used to calculate selection indexes to support breeders in multi-traits selection. A total of 6 indexes have been developed following the work of Quinton (2012) but 3 main indexes are used for 3 main purposes : terminal, maternal and higher prolificacy. The relatives weight of the traits within these 3 indexes are described below. For example, breeders having terminal breeds as Suffolk and Hamphire are recommended to use the Carcass index to improve growth and carcass traits, prolific breeds as Rideau Arcott and Romanov can use the Maternal index for a focus on mothering abilities whereas maternal breeds as Dorset and Polypay will have a focus on higher prolificacy with the Higher prolificacy index. The selection indexes are published for all animals of all the breeds which allows the breeders to select its sheep for all the purpose.

Results and discussion

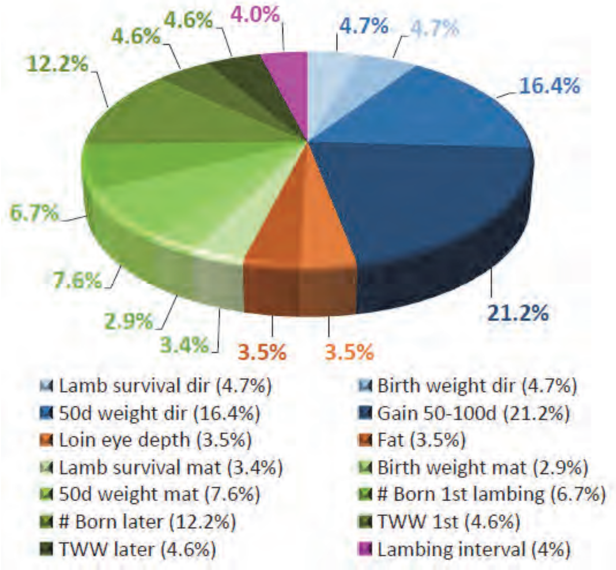
Over the past ten years, GenOvis has seen its number of members increase with the arrival of Ontario in the program (previously known as SFIP) in 2011 and two other industry organizations (OSF and CSBA) in 2016. Also, lots of effort were put to make producers aware of the importance of the records assiduity and consistency into the database. Since 2016, a 36 to 50% increase is observed in the number of records for growth traits, namely the number of lambs, the birth weight, the 50-day weight, and the 100-day weight (Figure 1A). Records for number of lambs are higher than every other growth trait (Figure 1A). Although the number of records is noticeably lower for loin depth, an important increase of ? 140% in its records has been observed over the past ten years (Figure 1A). The low records of loin depth compared to other growth traits are explained by the fact that this trait needs to be measured by ultrasound, thereby only a few producers pay for it and accredited technicians are not available in all provinces.



Carcass Index



Maternal Index



Maternal Higher Prolificacy index.

The number of lambs born per breed is quite stable in the past ten years for the majority of breeds, except, Crossbred, Romanov and Rideau Arcott (Figure 1B). Romanov and Rideau Arcott breeds have seen a 100% and XXX% increase in their number of lambs from 2016 to 2021 given the need to improve prolificacy for the producers. Crossbred group has a 50% increase in its number of lambs from 2014 to 2021 (Figure 1B) that could be explained by commercial producers being interested in recording data to follow their performance, do benchmarking and using breeding values for in-flock replacement. Many Crossbred lambs are hybrid or F1 lambs or produced by hybrid ewes as DP/RV for use in commercial flocks as replacement ewes or to produce market lambs.

The EBV 50-day weight genetic trend has been upward over the past 10 years, predominantly for Suffolk and Hampshire which are terminal breeds (Figure 2A). Although Romanov breed is a prolific breed with a slower growth rate, an interesting increase is seen (Figure 2A). For Dorset, Rideau Arcott, and Polypay breeds the increase is slower (Figure 2A). The same upward tendency is seen for the EBV gain 50-100d genetic trend (Figure 2B). Suffolk breed is still higher than any other breed. Polypay breed still have a slower increase for the EBV gain 50-100d genetic trend, and Romanov breed have again a good upward trend even if below zero (Figure 2B). The drops in the trend for the Dorset breed in 2017 can be explained by the sell of an important Dorset flock used for RandD purpose.

A slow upward trend is seen for the EBV fat depth genetic trend in Dorset, Rideau, and Polypay breeds (Figure 3A). After an upward trend, Hampshire breed had an important drop in 2019 for EBV fat depth to reach the same EBV seen in 2012 (Figure 3A). Since 2018, Suffolk breed had an impressive increase in its EBV fat depth (Figure 3A). Romanov breed, for its part, has a low downward genetic trend for its EBV fat depth (Figure 3A).

In the past 10 years, a rapid upward genetic trend of EBV loin depth is seen for Suffolk and Hampshire breeds that follow each other closely with an EBV around 2, in 2021 (Figure 3B). Since 2017, Dorset and Polypay breeds have seen an observable increase in their EBV loin depth genetic trend (Figure 3B). For Rideau and Romanov breeds, the EBV loin depth genetic trend is slowly increasing. The genetics correlations, breed purpose, the selection indexes used within each breed and the number of ultrasound measurement within each breed can explained partly these trends for carcass traits. Polypay, Rideau Arcott and Romanov breeds have adopted more recently the ultrasound measurements and for a small proportion of the population which are shower more recent significant improvement of loin depth. These breeds are also maternal or prolific breeds. Suffolk and Hamphire breeds are terminal breeds with a higher emphasis on carcass traits in the Carcass selection index which can be seen on the genetic trend of

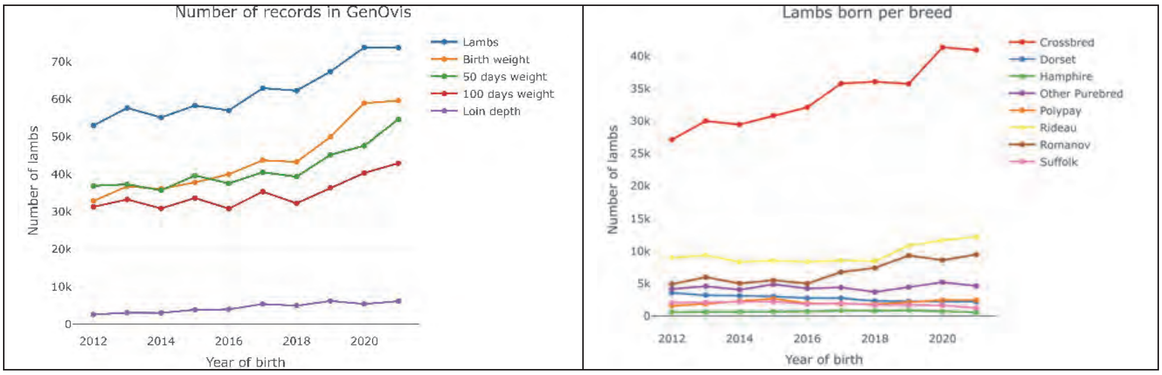


Figure 1. Records and breeds in GenOvis.

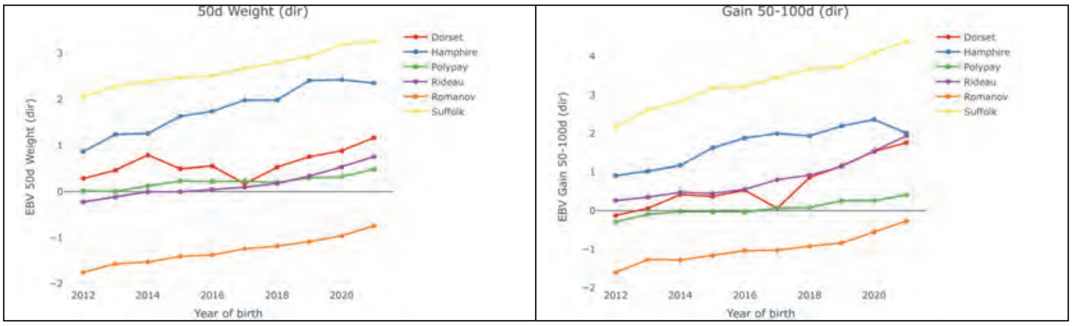


Figure 2. Genetic trends of growth traits per breed.

loin depth. Dorset is a mixture of being selected as a dual breed by some producers and having historically ultrasound measurement in their population. The fat depth is having a negative value in the selection indexes but is also unfavourably correlated genetically with loin depth and growth traits. The more recent increase of fat depth in the Suffolk breeds underline the relevance of taking ultrasound measurement and using the Carcass index.

The genetic trends of lamb survival are showing no increase for the Romanov and Dorset breeds and a slow increase for the other breeds. The direct EBV of Lamb survival is having a higher relative weight in the carcass index than the maternal and the higher prolificacy. However, some Suffolk and Rideau Arcott breeders seem to be more concerned about this trait and adding some value to it.

The Suffolk and Hamphire breeds are showing no improvement of the maternal 50 days weight. This trait is not included in the Carcass index but some cautious should be taken in the Hamphire breed in the coming years considering the decrease in 2021. The genetic trends are positive for the Polypay, Dorset, Rideau Arcott and Romanov and in line with the relative weight of this traits in the maternal and higher prolificacy indexes.

No specific trends are observable for Number of lambs born except a very slow improvement in the Dorset in more recent year. Again, it is consistent with the low heritability of the traits and the higher emphasis of the trait in the higher prolificacy index for the Dorset breed. A potential explanation for this newly trend in Dorset can be benefit from additional data coming from commercial flocks using crossbred ewes as Dorset/Romanov having many litters record for known Dorset sires. The genetic

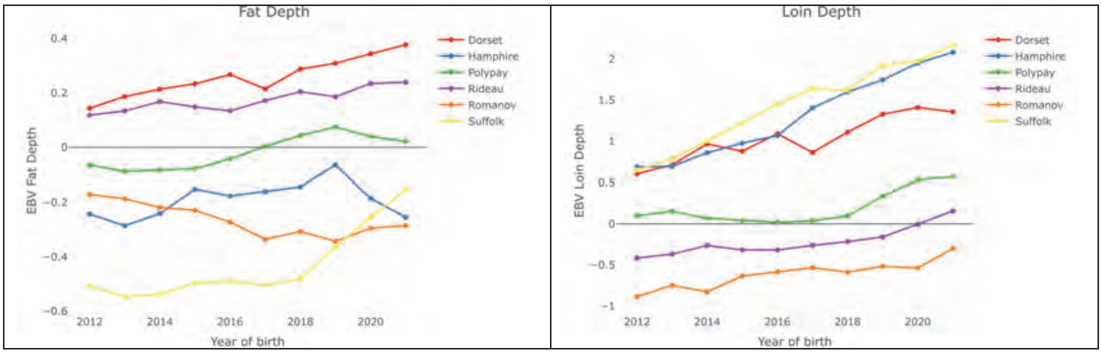


Figure 3. Genetic trends of carcass traits per breed.

trends of Total weight at weaning seem to increase more constantly in Romanov and Rideau Arcott breeds which are using the maternal index as selection index. However, the relative weight of this trait in the maternal and higher prolificacy indexes as well as the low heritability may explain the low trends and some inconsistencies.

In general, the average inbreeding levels of lambs born from 2012 to 2021 are showing slow increases. The change in inbreeding levels varied from negative in Polypay (-0.2%) to the largest positive increase in Dorset (+1.1%). For most of the breeds, inbreeding level trends can be acceptable considering the low increase per generation and the possibility to import breeding stock from other countries. The Rideau Arcott had some significant increase in their inbreeding levels until 2018 and some cautious has been taken in using mating plans to maintain it at a more constant level. This breed has been created entirely in Canada and cannot rely on breeding stock import.

Conclusion

The sheep meat sector is relatively small but important for the Canadian agriculture industry. More developments and innovation are needed and encouraging initiatives can be observed through better integration of genetic services in small ruminant and better standardization of data recording. GenOvis database is a gold mine of information that allows breeders to make genetic gains and to help producers make decisions for their flocks and improve their profitability. As the demand for lamb meat is on the rise in the past years, these precious data can be much more enhenced. Precision livestock farming (PLF) is a popular current concept that uses « big data » to make a more sustainable agriculture (Koltes *et al*., 2019). PLF is associated to different advantages as to improve precision of breeding, feeding, and animal health, productivity, and well-being (Guarino *et al*., 2017, Wolfert *et al*., 2017; Weersink *et al*., 2018). GenOvis data can be valued in technologies combining phenotypes and genomics and offers an opportunity to bring the PLF concept to the Canadian sheep industry. The association of phenotypic data to genomics allows the development of genomic breeding values and create genomic signature with genetic markers (ex.: SNP, CNV, etc.) specific to some economic traits of interest (Xie and Tammi, 2009; Uitterlinden, 2016). GenOvis offers vast opportunities for the Canadian sheep industry to address current and emerging challenges as economic viability, environmental sustainability and resiliency.

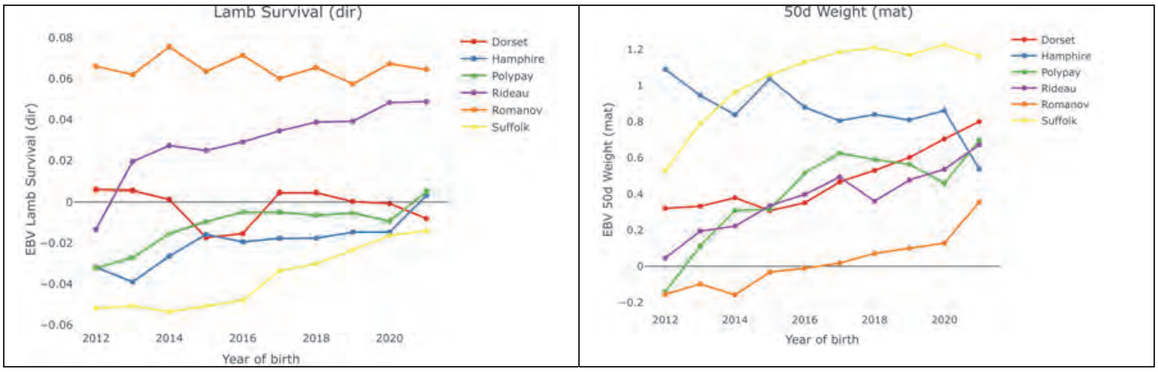


Figure 4. Genetic trends of lamb survival (dir) and maternal 50d weights per breed.

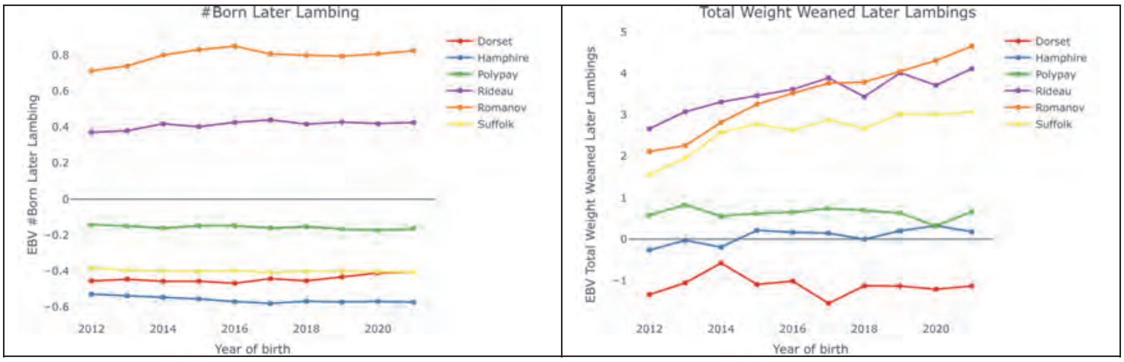


Figure 5: Genetic trends of number of lambs born and total weaning weight per breed

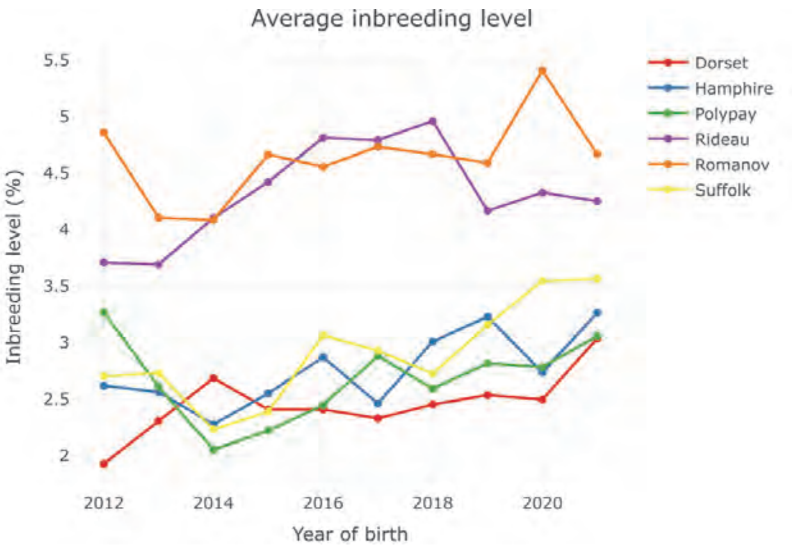


Figure 6. Average inbreeding levels (%) per breed.

References

Agriculture Canada. 2021. Customized report service – Market Overview – Canadian lamb and sheep meat and related product trends. https://agriculture.canada.ca/en/canadas-agriculture-sectors/animal-industry/red-meat-and-livestock-market-information/customized-report-service-market-overview-canadian-lamb-and-sheep-meat-and-related-product-trends (Consulted may 21st 2022).

Fortier, M.-P. 2022. Guide d'accréditation des mesures aux ultrasons chez l'ovin. Centre de développement du porc du Québec inc., Bibliothèque et Archives nationales du Québec. 23 p. ISBN 978-2-925175-02-5.

Animal Identification – New developments and future possibilities

D. Smith and K. Lane

Ceres Tag Limited, P.O. Box 552, Samford, Queensland, Australia
Corresponding Authors: david.smith@cerestag.com and kim.lane@cerestag.com

Abstract

Recently the world has been challenged by epidemics and most recently a human pandemic. This has disrupted the global economy and caused heartache for millions of people. Through this period of uncertainty, new ways to track people were developed using a carried smart phone that enabled check-ins, registrations, and proximity alerts. The threats today and of the future are the same for animals and an improved system(s) is required now.

Animals roam remotely and gather intensely; most animals though are where traditional communications are absent. They can't carry a device or recharge a smart phone. Yet globally, challenges of disease biosecurity and consumer traceability demand is intensifying. Therefore, the identification and traceability technology of the future needs to be dual operated in unlimited range and intense environments, reliable, automated, auditable, ubiquitous, interoperable, secure high integrity data, recording real time location and activity plus adaptable between intensive and extensive operations. The technology of now and the future should go further and record their impact on the ecology, environment, feed efficiency for genetic traits and health/welfare to determine if changes are required through management practices or other means to meet ESG expectations of the future from the supply chain to the consumer. The technology of now and the future should also incorporate one touch recording of transfers, providing information for the entire life of the animal.

However, there has been billions spent around the world on existing RFID systems. Therefore, if we disrupt and cause radical change too fast we will leave those investments obsolete or unable to serve out their lifetime. Instead, a transformative future technology needs to be introduced that is interoperable with existing infrastructure while rewarding those transferring to a new automated remote way of gathering information that is beyond an RFID number and includes real time behaviour/activity, location and proximity history of the animals during their lifetime. Transformative commercial technologies exist, and the approval processes are already dictated by international standards like ISO and international communication protocol requirements. Approval should only consider information collected and the way in which the technology interacts with the animal for ethical and welfare consideration. Long national approval processes should be removed and instead a centralised approval from a single body (ICAR) to assess so that these platforms may transparently operate to democratise the information about the food consumed globally.

Innovation is fast outpacing regulation. An international framework of global requirements to enable future rapid adoption of improved technologies for identification, traceability and activity are required now or risk substandard renegade schemes and technologies that operate outside of required biosecurity needs.

Keywords: Identification, traceability, transformative, interoperable, automated, biosecurity, regulation, standards, technology, ethical, animal welfare, democratise, international.

Introduction

Today we are faced with the frequency of biosecurity threats at an increasing rate. The current systems, regulations and standards implemented by nations, however proud of what they have, are insufficient to cope with biosecurity incursions of now and the future. We will see new biosecurity devastation from diseases, viruses and bugs that we had never previously known about or had been laxed to their incursion. Nations must be flexible and allow for new processes to be implemented quickly to cope with an ever-changing dynamic that had evolved as the world has become more global and movements more frequent between borders. Even then, as climate change continues to evolve, we will see diseases cross borders through insects or through airborne movement. These threats are very real with many current and recent examples of how quickly things may evolve. Nations can no longer put their head in the sand and pretend that the systems they have today will hold them in good stead for generations. There are many examples where generational change is necessary to implement new ways of doing things, new technologies and new frameworks. We can not afford to sit on our laurels of past and hope for the best.

Double this requirement when considering the consumer who has never known less but wants to know more about where their food comes from for provenance, animal welfare, sustainability, food security and safety.

So how do we achieve all of these demands without stifling the food production system through regulation and outdated systems and processes?

There has been very few, if any new types of identification traceability systems implemented for decades in any country. Those nations that implemented RFID systems have led the world for traceability for decades however they are now struggling with the process changes required when considering the new environment in which the world operates. They stand by what they have previously achieved and a mindset that “if it ain’t broke, don’t fix it”, completely impervious to what is required now and in the future. Those nations set the benchmark and toiled to gain acceptance of the system at the time that satisfied the limited need of the past and subsequently are unable to shift mindset that those capabilities are now insufficient and outdated to today’s desperate requirements.

Ironically, it is nations without current identification systems that have the opportunity to implement a fresh mindset and leapfrog the current leaders for world leading traceability and competitive advantage. These nations and those willing to accept that the status quo is insufficient and requires change, will be the new leaders in food security and food safety and create a distinct competitive advantage.

New dynamic frameworks with centralised assessment and clear interpretation of what must be achieved from any technology must be available to ensure that regulation keeps pace with innovation. These frameworks become less how things are done but rather the parameters they provide to achieve the needs of the changing environment. These new technologies or processes should also be transparent enabling national systems and/or consumers to access if desired or required. This type of thinking is novel for regulators who have previously had success with outdated systems and who are reluctant to change to meet the demands of today's producers and consumers to automate and democratise information critical to the supply of food and associated products to an ever growing and more demanding global population.

Technologies

Technologies are evolving that can do many things and to prescribe a single methodology is both limiting and perhaps careless. Today and in the future, we face challenges of speed to face biosecurity challenges, identification traceability and consumer demands.

There are a range of technologies currently available from facial recognition, RF frequency identifiers, bolus, injectables, drones, satellite trackers and other types of markers and monitoring. Most of these have prescribed purposes and require infrastructure to be established to operate making them expensive to implement, manual to operate and limited to particular purposes.

Regardless, this should not limit their applicability if they have the ability to reliably provide the information prescribed at every stage of the supply chain or if an animal was to only stay at the particular stage of the supply chain and never leave where that technology is able to be applied.

Scalability remains the primary challenge of most technologies and the ability to implement with existing systems while transitional activities occur. That is, the next stage of provenance proof will need to be transformative incorporating existing ways of identification traceability while building future automated multi-times daily reporting systems to display and record insights of the animal during its lifetime.

The operations of animals during their lifetime also transition between extensive operations, often required for breeding, sustainability, and animal welfare purposes to the intensive food production environments. Technologies that operate in one environment, often do not operate in the other creating a limiting effect to their usability in a supply chain.

While software systems are good for management, visualisation and blockchain, they do not create the data from which those platforms are populated. They are necessary but by themselves, not a solution.

Currently there is only one such transformative technology that incorporates satellite technology that can transition to Bluetooth that is interoperable in an intensive and extensive environment providing identification traceability and biosecurity contact tracing capabilities as well as performance monitoring. While others excel in their particular field, the adoption rate will be limited making it difficult to prescribe in national cases unless every sector of the supply chain is considered. Other technologies only target a sector of the market however large or small, and require additional input or technologies to transition data from one platform to another and most operate a closed rather than interoperable ecosystem.

ICAR and National Traceability Systems

ICAR provides the guidelines that some nations adopt as their national requirement to operate while others prescribe additional requirements that take many years, are not practical nor are they consistently repeatable and appear to be a block for any transformational or improved technologies due to potential commercial changes and replacements. The latter are usually nations with existing standards that find it difficult to improve from methodologies implemented decades ago with incumbent suppliers who are very well known, comfortable with the status quo. The reason to change often receives pushback for commercial preferences rather than for the betterment of the industry. Some individual national programs, however great or small in their differences, have resistance towards change as individual interests of committees and sub committees can take precedent over the intended good for systems that were implemented decades ago which may no longer be able to manage the threats and purpose of today.

The continuation of resistance to change or to adopt improved technologies that increase speed to market while providing increased quality control, only increase the likelihood of renegade new platforms to enter the space, this has happened in other industries such as Social Media platforms to traditional media or rideshare to the taxi industry.

The guidelines for traceability, especially for exporting markets, should be flexible enough to incorporate new technologies and governed by principals of a central body assuring which ever technology is able to provide the necessary information needs. This would increase the international competitiveness for supply of animal products and ensure that the required information is able to be supplied in an automated timely nature with minimal input required from the producers themselves. The information should be democratised, automated, fluid and transparent for regulators and consumers to see.

Consequences of not adopting new technologies

Adoption of new technologies also increases the rate of digital transformation so that storage for historical and traceability purposes can be done in a reliable, efficient, fast, compact and cost effective way. Currently so many national traceability systems are paper based with digital options. The paper systems are where information is lost, unreadable or sometimes frauded. Existing systems are often manipulated, only inserting tags at the final moment between movements with no way of knowing where the animal originated or any of its history.

Biosecurity threats and incursions are coming more into focus as international movement via ships, planes and road based transport systems that carry animals or animal products which may be carrying diseases or parasites across borders into other nations or regions. When this occurs, at best we currently only have passive systems to rely on to provide the information required and this is only valid if all of the paper based systems were correctly completed, submitted and considered. This reliance on humans to do the right thing or do things correctly every time in every place is a massive risk to the world and our ability to feed a growing population safely and reliably.

Only automated technologies that can demonstrate the ability to operate without any additional infrastructure and with the capability to report across borders can be considered for any serious implementation at scale or in nations where export is part of their drivers for production.

Never before have we seen such significant threats take hold around the world and cause such devastation to the global food supply. Examples are African Swine Fever where approximately 50% of Chinas' swine supply was decimated, Mycoplasma Bovis

Biosecurity threats

in New Zealand or currently the Foot and Mouth and Lumpy Skin Disease in Indonesia threatening to jump the border into Australia, one of the worlds largest exporters of beef products.

The damage done from these incursions reaches into the billions of dollars, lost market access and reputation plus a major deployment of resources to manage the situation, placing people, eco-systems and animals at risk for no economic or social gain.

The intense systems are not likely to escape these incursions in many supply chain systems with breeding in extensive operations and then feeding into the intense operations where a single infected animal can cause hundreds or thousands to be infected in a relatively small space and short timeframe. The animals in the extensive production systems are also exposed to wild animals that may be carrying the diseases and are able to move beyond boundaries providing the interconnectivity to spread from one animal to another. The movement along supply chains will also spread biosecurity incursions rapidly, with no illegality intended, often operating with the innocence of those who do it as their way of life, without full engagement of inspection and often in remote locations. For this reason, we need to have notifications both on the animal that can alert through an LED or similar and to alert digitally when an event occurs for the early detection and then subsequent contact tracing to determine the extent to which the incursion has occurred and its origins, most likely through genomic sequencing.

Conclusion

The deployment of Animal Identification technologies to support digital transformation cannot be ignored, swept away or pretended not to exist. They are real and available now in multiple countries. The current traceability systems are not robust enough to protect for tomorrow's challenges and provide the information required for protection or to inform a growing inquisitive population on where their food comes from. International frameworks must be derived and deployed as soon as possible so that regulation may keep pace with innovation, or the market will create their own principles and acceptances to operate seeking forgiveness rather than permissions that are irrelevant to today's needs and opportunities.

There are technologies that can provide all of the needs of the future with unlimited range and machine learning to continually update to the needs of tomorrow, that is scalable, reliable, interoperable and critical to the future of the animal recording. These technologies will have others join them in capability in the future to ensure they also have competitiveness and challenge to continually improve.

New frameworks are required now, technology development is moving at a rapid pace. If agencies and nations are not ready, producers will rely on ways to get their products to market. The industry is undergoing a significant demographic shift and the future generations will not wait. Allowances and guidelines that speak to outcomes rather than methodologies are required if we are to support digital transformation across livestock production and build a prosperous and effective safe food future to feed a rapidly growing population.

Acknowledgement

Meat and Livestock Australia - serving red meat and livestock producers | Meat and Livestock Australia (https://www.mla.com.au/).

Home | Integrity Systems (https://www.integritysystems.com.au/)

Commonwealth Scientific and Industrial Research Organisation, Australian Government - CSIRO (https://www.csiro.au/)

Queensland Government | Queensland Government (www.qld.gov.au)

Die Identitas - Identitas (https://www.identitas.ch/)

Ceres Tag (https://www.cerestag.com/)

Satellite Solutions and Services | Globalstar US (https://www.globalstar.com/en-us/)

Double Injection Molding - Yomura Plastic Solutions Provider (https://www.yomura.com.tw/)

Rezare Systems (https://www.rezare.co.nz/)

Product Design Brisbane | Clandestine Design Group (https://www.clandestinedesigngroup.com/)

ADE implementation at iDDEN

J. Frandsen

SEGES Innovation, Denmark

Abstract

Exchange of data between on farm equipment and farmer owned Dairy Data organization (DDO) databases has been a subject for years. With the updated ADE standard and the iDDEN integration we see the following advances to be ready to be harvested for the different stakeholders

- MRO (milk recording organization) / DDO (Dairy Data organization)
  - Data for milk recording
  - Data for breeding evaluation
  - Data for decision support
  - Only one integration (there will be some minor extra work pr OEM)
- OEM (management system in connection to on farm equipment)
  - Only one integration (there will be some minor extra work pr MRO/DDO)
  - Maybe even better market position
  - Data back up
- Farmers
  - Data only have to be recorded once
  - Better data - Les bugs in data
  - Better decision support on more data from management systems
  - Easier daily routines – les trouble

Introduction

Exchange of data within the dairy business has been a subject for many years. The farmer and the service sector around the farmer see a growing need to have data available wherever relevant. With the dramatical growing number of sensors in the business and the increasing number of data the challenge is only growing

Throughout the years the Electronic Data exchange, ADE, working group in ICAR has been working on a harmonized data description and some entities in the market has started the implementation of the standard. One of them is iDDEN, who is a not profit company of DDOs from the countries listed in Table 1

Table 1. List of countries participating in the ICAR ADE WG.

- Shareholders
  - CRV
  - DataGene
  - Lactanet
  - NDHIA
  - NCDX ApS
  - RDV
  - vit

Designated Area Responsibilities

- The Netherlands & Belgium
- Australia
- Canada
- USA
- Den, Ice, Fin, Nor, Swe
- Austria, Germany
- Germany, Luxemburg

With the whole set up in place we are see that we are about to overcome the challenge about “who to be the first if nobody else are using the system”

Experiences of implementation

The Danish DDO (Dairy Data Organization) owned by the farmers organization and operated by SEGES innovation has for more than 10 years work with exchange of data first on national level, then in cooperation with the other Nordic organisation, where the NCDX system was developed and latest in a worldwide perspective with the iDDEN system. The main challenge has been that all supplies of on farm equipment must integrate to a variety of different DDOs and wise versa seen from DDOs point of view – in the perspective that the farmer is asking for the exchange

With development of NCDX we saw positive attitudes in the market, and some integrations that increased the amount of data which were exchanged. When iDDEN was established and used the NCDX as basic and decided to upgrade to the ADE standard it seems like a breakthrough. There is a growing interest from suppliers of equipment and DDOs outside iDDEN

Which data

From the start there has been a demand to start with most used data like reproduction data. Later came other data items like milking data, health data, which are the main categories in the iDDEN system today. The wishlist for more data items is long and are dynamic prioritized by the ADE WG and the users of the ADE standards. Next in line is right now sensor data, feeding and group data

Data both way

To cover the variety in usages and practice in different situations, it has been important for iDDEN that data can be exchanged both ways. In the most ideal situation this means that the farmer can choose where to enter data and that data are available where needed

Varity in which data to exchange

In the iDDEN agreement is included a data exchange agreement which is meant to be local. The data exchange agreement will regulate which data are to be exchanged by the individual supplier of equipment and the individual DDO. In that way it is possible to respect the different business situation, there is in different local areas

Status to day

From the DDOs in iDDEN 4 have been working on there integration and on the supplier side at least to are working on the integration, and more are about to start. In Denmark we expect to go in production test on a farm and test on real data in June 2022

Conclusion

The challenge of getting data to flow between the stakeholders around the farmer can be solved. By having an independent body as the ADE WG in ICAR, and non-profit company to take care of the implementation in an IT system, it is possible to develop and maintain a HUB so cost effective that it is attractive for the different stakeholders. We hope this will continue with increasing user, because with more users the fee for using the system will decrease

Acknowledgement

Thanks to iDDEN group, both the shareholder representatives and other related who have worked on making iDDEN a valuable system for the farmers that are listed below:

CRV (Netherlands & northern Belgium)	Bert van't Land	+31 653907474	Bert.van.t.Land@crv4all.com
DataGene (Australia)	Matthew Shaffer	+61 438 761 822	mshaffer@datagene.com.au
Lactanet (Canada)	Neil Petreny	+1 519 824 2320	npetreny@lactanet.ca
National Dairy Herd Information Association (USA)	Jay Mattison	+1 608 848 6455	jmattison@dhia.org
NCDX (Nordic Countries)	Johannes Frandsen	+45 2171 7720	jhf@seges.dk
RDV (Austria & Germany)	Florian Grandl	+49 89 544 348 926	Florian.Grandl@lkv.bayern.de
vit (Germany)	Reinhard Reents	+49 4231 955 140	reinhard.reents@vit.de

Can milk analysis by infrared spectroscopy reveal the welfare status of cows?

M. Bahadi[1,2], D. Warner[2], A. A. Ismail[1], D. E. Santschi[2], D. M. Lefebvre[2],
R. Duggavathi[1] and E. Vasseur[1]

[1]McGill University, 21111 Lakeshore Road, H9X 3V9,
Sainte Anne de Bellevue, QC, Canada
Corresponding Author: mazen.bahadi@mail.mcgill.ca
[2]Lactanet, 555 Boul des Anciens-Combattants, H9X 3R4,
Sainte Anne de Bellevue, QC, Canada

Abstract

Assessing welfare status of cows currently requires farm visits. However, detecting welfare status of cows using unbiased analyses of milk samples is a practical alternative as millions of milk samples are routinely collected for milk recording. Thus, we set out to isolate milk infrared spectral fingerprints representing effects of housing-configuration modification on milk composition that may be associated with changes in welfare status of cows. We applied a new analysis method to milk spectra by combining principal component analysis and mixed modelling. This new method was used in trials examining the impact of housing configurations, including 4 tie rail positions and 2 chain lengths, on milk composition. Principal components extracted from the averages of milk spectra collected during weeks 8-10 of the trials revealed a significant effect of the housing configuration treatment on those spectra. This analysis method was capable of capturing changes in milk composition that were attributed to negative and positive changes in welfare status of cows resulting from housing modifications to tie rail positions and chain length, respectively. Spectral analysis revealed higher levels of biomarkers related to body fat mobilization (i.e., beta-hydroxybutyric acid, acetone, and citrate), in milk from cows subjected to the tie rail position that had most restrictive access to feed which may have resulted in a possible reduced feed intake (not measured in the trial) that led to an elevated body fat mobilization. No changes in energy reserves estimable using visual body condition scoring were observed for those cows. These results were corroborated by increased injuries at two locations on cows' necks because of pressure on their neck through repeated contact with the tie-rail, which may be indicative of laboured access to reach feed. In addition, milk from cows tied with longer chains had lower levels of biomarkers linked to episodes of ruminal acidosis (i.e., milk non-protein nitrogen and trans fatty acids). Behavioural observations showed that these cows spent more time with their heads in the manger area, assuming they might have chewed more; hence, they might have produced more saliva to balance the ruminal pH. To conclude, this novel spectral analysis methodology offers a new tool for assessing cow welfare status by detecting trends of changes in milk composition in their early stages, which will provide a mean for remote and unbiased detection of cows or herds with welfare problems before appearance of clinical signs.

Keywords: FTIR spectroscopy, cow welfare, housing configuration.

Introduction

Assessing welfare status of cows currently requires farm visits. However, detecting welfare status of cows using unbiased analyses of milk samples is a practical alternative as millions of milk samples are routinely collected for milk recording. Several studies have demonstrated that milk composition reflects the concentrations of key blood plasma metabolites, such as non-esterified fatty acids (NEFA) (Jorjong *et al*., 2014) and beta-hydroxybutyric acid (BHB) (Pralle *et al*., 2018), the nutritional state of the cow (Weber *et al*., 2013, McParland *et al*., 2014) and health conditions (Arnould *et al*., 2013) that might affect the cow's productivity. Other studies have demonstrated that health issues can be associated with behavioural changes that precede the clinical diagnosis. For example, cows that were diagnosed with a left displaced abomasum showed increased step activity in comparison to healthy animals during the week prior to the clinical diagnosis (Petersson-Wolfe *et al*., 2017). Cows at risk of subclinical ketosis were observed to have fewer feeding visits and 18% less dry matter intake in comparison with healthy controls in the week leading up to calving (Goldhawk *et al*., 2009). It was also observed that lying time was greater and average daily steps were significantly lower five days before cows were diagnosed with mastitis (Yeiser, 2011), and 2 days before onset, lying time decreased compared with that of healthy cows (Petersson-Wolfe *et al*., 2017). In this study, we hypothesized that cow welfare status might lead to physiological changes, which might be reflected in milk chemical composition. Hence, we set out to isolate milk Fourier transform infrared (FTIR) spectral fingerprints representing effects of housing-configuration modification on milk composition that may be associated with changes in welfare status of cows. We applied a new analysis method to milk spectra, which was described in a previous publication (Bahadi *et al*., 2021).

Materials and methods

The animal trials were conducted at the Macdonald Campus Dairy Complex of McGill University (Sainte-Anne-de-Bellevue, QC, Canada).

Trial 1: tie-rail configuration

Forty-eight lactating Holstein cows were assigned to 4 tie-rail (TR) configurations varying in height and position (Table 1). Details about these treatments can be found elsewhere (St John *et al*., 2021). Cows were assigned to 6 blocks to account for parity (primiparous: n = 12, multiparous: n = 36), days in milk (DIM; early: 0–100 d, mid: 101–200 d or late: 201–305 d), and location in the barn prior to the start of the experiment. Cows were housed in two separate rows of tie-stalls facing the barn wall and they were in trial for 10 weeks with 24 cows starting in summer 2016 (period 1: from July 25th to October 3rd) and the remaining 24 cows starting in fall 2016 (period 2: from October 10th to December 19th).

Trial 2: chain length configuration

Twenty-four lactating Holstein cows were assigned to 2 tie chain length (TCL) treatments (Table 1). Details about these treatments can be found elsewhere (Boyer *et al*., 2021). Cows were assigned to 12 different blocks of two cows to account for age of the cow (i.e., parity or number of lactations) and days in milk within current lactation (average DIM 129) and were placed evenly in two rows facing a wall within the barn. The trial lasted for 10 weeks from February 20th, 2017, to May 1st, 2017.

Milk samples

One composite milk sample per week was collected from each cow participating in the trials. The sample consisted of milk collected during the evening milking and the morning milking of the next day. All collected milk samples were analysed for milk composition by FTIR spectroscopy at the Lactanet laboratory (Sainte-Anne-de-Bellevue, QC, Canada) using the same CombiFoss FT+ analyser (FOSS, Hillerød, Denmark).

Spectral analysis

Baseline and treatment application average spectra were calculated for each cow from spectra of samples collected from weeks 1 to 3 and from weeks 8 to 10, respectively. To detect the housing treatment effect, we applied a new analysis method to those spectra by combining principal component analysis (PCA) and mixed modelling. Details of this methodology are described elsewhere (Bahadi *et al*., 2021) and are summarized in Figure 1. The statistical models that were used in the analysis were:

For the TR trial

$$Y_{ijk} = \mu + trt_i + period_j + block_{kji} + e_{ijk} \qquad (1)$$

where trt_i was the fixed effect of the i^{th} TR configuration treatment, $period_j$ was the fixed effect of the j^{th} period, $block_{kji}$ was the fixed effect of k^{th} parity, DIM and location in the barn from the period on the TR configuration treatment and was the random residual error.

For the TCL trial

$$Y_{ijkl} = \mu + trt_i + row_j + block_k + e_{ijk} \qquad (2)$$

where trt_i was the fixed effect of the i^{th} TCL treatment, $block_k$ was the fixed effect of the k^{th} parity and lactation stage combination, row_j was the random effect of the j^{th} row in the barn and e_{ijk} was the random residual error.

Results and discussion

Principal components (PC) 7 and 6 revealed significant housing treatment effects for the TR ($P = 0.011$) and TCL ($P = 0.032$) trials, respectively (Table 2). Both PCs did not reveal significant effects for the other studied factors that were included in their respective statistical models. This observation suggests that the TR configuration and TCL might have influenced milk composition during these two trials. Differences in the least-squares means for scores of PC7 and PC6 for the TR and TCL trials, respectively, suggest that composition of milk from cows enrolled in T3 was significantly different from that of cows enrolled in T1 for the TR trial and the composition of milk from cows enrolled in T1 was significantly different from that of cows enrolled in T2 for the TCL trial (Table 3).

For the TR trial, inspection of the integral of the loading spectrum for PC7 revealed an inverse relationship between lactose and biomarkers related to body fat mobilization

Table 1. Housing configuration treatments across the 2 trials.

Trial	Housing configuration	Treatments			
		T1	T2	T3	T4
1	Tie rail[1]	122 cm x 36 cm	122 cm x 18 cm	112 cm x 18 cm	112 cm x 36 cm
2	Chain length	1.0 m	1.4 m	-	-

[1] Height from the stall base x forward position from the manger wall

Table 2. Principal components extracted from treatment application spectral datasets that revealed significant effects. The table also lists P values obtained from the SAS Mixed Procedure for tested effects in each trial.

Trials	Spectral Dataset[1]	PC	Eigenvalue	Explained Variation %	P Values		
Tie rail	VN FD	7	3.82	1.37	Trt[2] 0.011	Period 0.559	Block 0.060
Chain length	VN FD	6	4.75	1.70	Trt 0.032		Block 0.088

[1] FD = first derivative, VN = vector normalized
[2] Trt = treatment

Table 3. Differences of least squares means for the scores of the principal components that revealed significant treatment effect.

Trial	Trt[1]	Trt	Estimate	Standard Error	DF[2]	t Value	P Value	Scheffé Adj. P Value
Tie rail	T1	T3	2.1442	0.68	30	3.15	0.004	0.033
Chain length	T1	T2	-1.6819	0.6650	9	-2.53	0.032	0.032

[1] Trt = treatment
[2] DF = degrees of freedom

(i.e., BHB, citrate, acetone). This observation can be interpreted that cows subjected to the tie rail position T3 had the most restrictive access to feed which may have resulted in a possible reduced feed intake that led to an elevated body fat mobilization. No changes in energy reserves estimable using visual body condition scoring were observed for those cows (St John *et al*., 2021). The behavioural results (St John *et al*., 2021) showed that cows in T3 had increased injuries at two locations on cows' necks because of pressure on their neck through repeated contact with the tie-rail, which may be indicative of laboured access to reach feed.

For the TCL trial, inspection of the integral of the loading spectrum for PC6 revealed that milk from cows tied with longer chains had lower levels of biomarkers linked to episodes of ruminal acidosis (i.e., milk non-protein nitrogen and trans fatty acids). This observation suggests that cows with a longer chain may have a more stable ruminal pH, which is an indication of better digestive health. Behavioural observations showed that these cows spent more time with their heads in the manger area (Boyer *et al*., 2021). Increased access to the manger could mean easier access to feed, and by doing so, more time was available for rumination and chewing; hence, they might have produced more saliva to balance the ruminal pH.

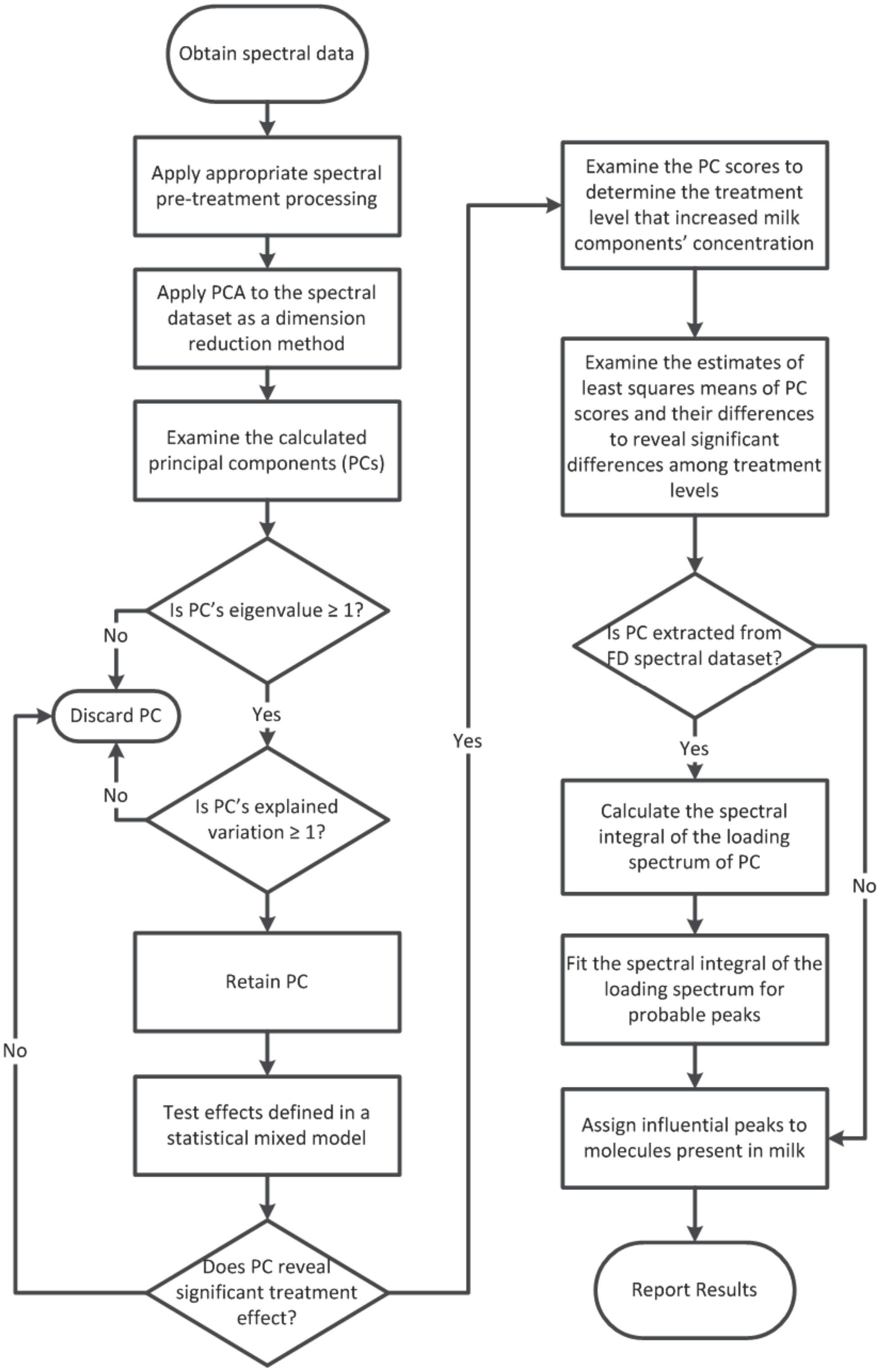


Figure 1. Workflow of hybrid data analysis approach of FTIR milk spectral data that combines multivariate analysis with mixed modelling to detect housing treatment effect (Bahadi et al., 2021).

Conclusion

Milk FTIR spectra could detect changes in milk composition that could be attributed to negative and positive welfare status in cows. Behavioural and clinical results corroborated the findings of the new spectral analysis methodology. This methodology offers a new tool for assessing cow welfare by detecting trends of changes in milk composition in their early stages, which will provide a mean for remote and unbiased detection of cows or herds with welfare problems before appearance of clinical signs.

Acknowledgement

Part of the material presented in this conference paper has been published in peer-reviewed scientific paper Bahadi *et al*. (2021) and is currently in review for publication in *Journal of Dairy Science* (Bahadi *et al*., manuscript ID JDS.2022-22042).

Jessica St John (McGill University) conducted the tie rail trial and developed the SAS code to detect the treatment effect on milk components by the mixed procedure as part of her M. Sc. Thesis. Behavioural and other animal welfare outcomes for the tie-rail trial are published in St John *et al*. (2021).

Véronique Boyer (McGill University) conducted the chain length trial and developed the SAS code to detect the treatment effect on milk components by the mixed procedure as part of her M. Sc. Thesis. Behavioural and other animal welfare outcomes for the chain length trial are published in Boyer *et al*. (2021).

The animal trial was funded by Vasseur’s NSERC-Novalait-Dairy Farmers of Canada-Lactanet Industrial Research Chair on Sustainable Life of Dairy Cattle. Complementary funding for the study (physiological indicators) was provided by NSERC Collaborative Research Development Grant (Project no. CRDPJ 499520 16) and FRQNT-CRIBIQ-Novalait through Programme de recherche en partenariat pour l’innovation en production et transformation laitières (Project no. EPI 2016-138) and Lactanet.

The spectral analysis was performed at the McGill IR group.

References

Arnould, V. M.-R., R. Reding, J. Bormann, N. Gengler, and H. Soyeurt. 2013. Review: milk composition as management tool of sustainability. Biotechnologie, Agronomie, Société et Environnement 17(4):613-621.

Bahadi, M., A. A. Ismail, and E. Vasseur. 2021. Fourier Transform Infrared Spectroscopy as a Tool to Study Milk Composition Changes in Dairy Cows Attributed to Housing Modifications to Improve Animal Welfare. Foods 10(2):450.

Boyer, V., A. de Passillé, S. Adam, and E. Vasseur. 2021. Making tiestalls more comfortable: II. Increasing chain length to improve the ease of movement of dairy cows. Journal of Dairy Science 104(3):3316-3326.

Goldhawk, C., N. Chapinal, D. Veira, D. Weary, and M. Von Keyserlingk. 2009. Prepartum feeding behavior is an early indicator of subclinical ketosis. Journal of Dairy science 92(10):4971-4977.

Jorjong, S., v. A. Knegsel, J. Verwaeren, M. Val Lahoz, R. M. Bruckmaier, D. B. Baets, B. Kemp, and V. Fievez. 2014. Milk fatty acids as possible biomarkers to early diagnose elevated concentrations of blood plasma nonesterified fatty acids in dairy cows. Journal of Dairy Science 97(11):7054-7064.

McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. Odonovan, S. T. Butler, J. E. Pryce, and D. P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. JODS Journal of Dairy Science 97(9):5863-5871.

Petersson-Wolfe, C., N. Steele, T. Swartz, and B. Dela Rue. 2017. Opportunities for identifying animal health and well-being disorders using precision technologies. in Pages 1279–1292 in Large Dairy Herd Management. Am. Dairy Sci. Assoc., Champaign, IL.

Pralle, R. S., K. W. Weigel, and H. M. White. 2018. Predicting blood â-hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables with multiple linear regression, partial least squares regression, and artificial neural network. Journal of Dairy Science 101(5):4378-4387.

St John, J., J. Rushen, S. Adam, and E. Vasseur. 2021. Making tiestalls more comfortable: I. Adjusting tie-rail height and forward position to improve dairy cows' ability to rise and lie down. Journal of Dairy Science 104(3):3304-3315.

Weber, C., C. Hametner, A. Tuchscherer, B. Losand, E. Kanitz, W. Otten, S. P. Singh, R. M. Bruckmaier, F. Becker, W. Kanitz, and H. M. Hammon. 2013. Variation in fat mobilization during early lactation differently affects feed intake, body condition, and lipid and glucose metabolism in high-yielding dairy cows. Journal of Dairy Science 96(1):165-180.

Yeiser, E. E. 2011. The use of activity measures in combination with physiological factors as indicators of disease in dairy cattle. Virginia Tech.

Report on the incidence of selected hereditary disorders in the Polish population of Holstein-Friesian cattle

M. Gozdek[12]*, D. Kamola*[1]*, A. Prostek*[1] *and T. Sadkowski*[1]

[1]*Warsaw University of Life Sciences, Department of Physiological Sciences,*
Nowoursynowska 166, 02-787 Warsaw, Poland
Corresponding Author: m.gozdek@pfhb.pl
[2]*Polish Federation of Cattle Breeders and Dairy Farmers, Zurawia 22,*
00-515 Warsaw, Poland

Abstract

DUMPS (deficiency of uridine monophosphate synthase) and Haplotypes Holstein (HH1, HH3, HH4, HH5, HH6, HH7) are hereditary lethal autosomal recessive disorders that may affect Holstein cattle. The aim of the research was to determine the frequency of selected unfavorable mutations in the Polish population of Holstein-Friesian cattle. In the study 12 754 Polish Holstein-Friesian females were genetically tested. Results of the present study indicate, that the Polish population of dairy cattle is free from DUMPS. It turns out, that in 2019 the highest number of carriers were HH3 (5.81%), and the least HH7 (0.49%). In 2020, the highest number of carriers was HH5 (6.95%), and the least, similar to 2019, HH7 (0.45%).

Keywords: Cattle, gene frequency, microarrays, embryo mortality.

Introduction

Fertility is one of the most important traits in animal production. Significant decrease in fertility of dairy cattle in Poland has been arising for many years - a decrease in the fertilization and normal pregnancy rate is observed. These problems may result from the accumulation of genetic diseases in the population. Since individual sires generate tens of thousands of progeny via artificial insemination, cattle breeding populations are susceptible to the propagation of recessive diseases. The resulting threat was not immediately recognized because the heterozygous animals (i.e. asymptomatic carriers) do not express the symptoms of disease. Due to the frequent occurrence of genetic defects in combination with high breeding values, they tend to spread throughout the population. The increasing incidence of genetic diseases in cow herds is also the result of increased inbreeding within the entire world’s cattle population. There are several specific genetic disorders associated with Holstein cattle. Particular attention should be paid to genetic defects causing embryo mortality. Among the most important defects are HH1, HH3, HH4, HH5, HH6, HH7, and DUMPS (Table 1).

Material and methods

Genetic data used in the present study was collected in the process of routine estimating breeding value (EBV). Poland as a member of EuroGenomics Cooperative, uses their customized arrays. Material for genotyping (12 754 samples) was collected in the years 2019 to 2020. Ear punch samples were collected with the use of AllFlex Tissue Sampling Unit (TSU). DNA extraction was processed with the use of Clean BloodandTissue DNA Kit (CleanNA, Netherlands) according to the producer manual in KingFisher DUO DNA processor (Thermo Scientific, USA). Normalized samples were processed according to Illumina HTS protocol. Beadchips were immediately scanned on Illumina iScan system, scans were analysed using GenomeStudio Software.

Results

Results of the present study indicate that the Polish population of Holstein dairy cattle is free from DUMPS (since 1999 testing of Polish population of Holstein dairy breeding bulls is mandatory). It turns out that in 2019 the number of HH1 carriers was 2,95%, HH3 - 5,81%, HH4 – 2,33%, HH5 – 5,16%, HH6 – 1,83%, HH7 – 0,49% (Holstein-Friesian population). In 2020 the number of carriers was correspondingly HH1 3,11%, HH3 – 4,42%, HH4 – 1,51%, HH5 – 6,95%, HH6 – 1,94%, HH7 – 0,45% (Figure 1).

Discussion

No DUMPS carriers were detected in the present study. Our research results are also in accordance with the results of studies by Patel *et al.* (2006) and Oner *et al.* (2010) who reported no carriers respectively in Indian and Turkey dairy cattle populations. Similarly, Korkmaz Agaoglu *et al*. (2015), Koshchaev *et al.* (2018), Debnath *et al.* (2016) and Citek *et al.* (2006) did not detect any DUMPS carrier in their studies. The number of HH1 carriers in Poland is higher than in Brazil Albertino *et al.* (2022) and the USA

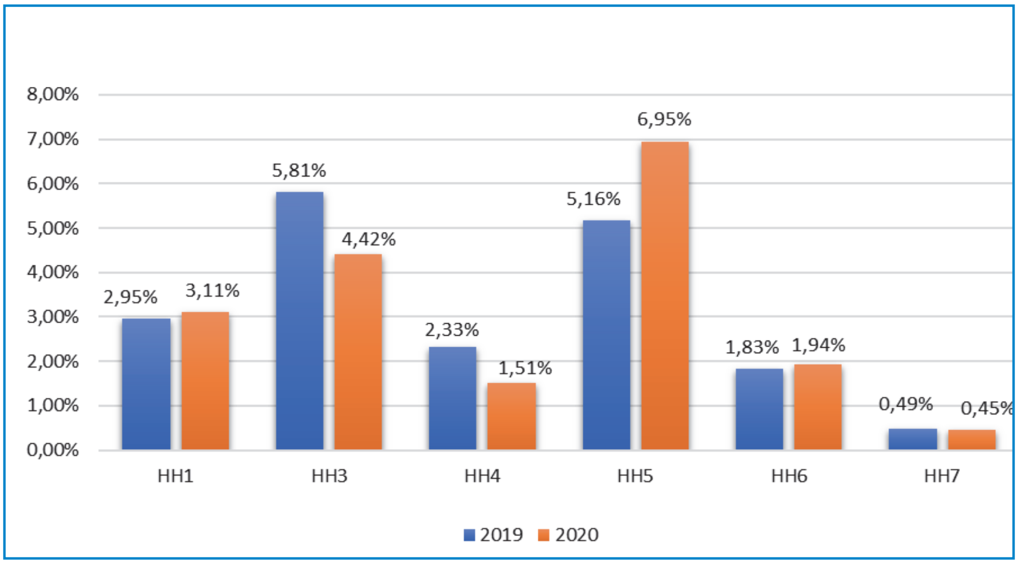


Figure 1. Percentage of carriers of selected hereditary disorders in the Polish population of Holstein-Friesian cattle.

Table 1. Basic characteristics of selected genetic disorders.

Disorder name	Species-specific name	Gene	Type of mutation	Year first reported
HH1	Haplotype HH1	APAF1	C>T	2012
HH3	Haplotype HH3	SMC2	T>C	2013
HH4	Haplotype HH4	GART	A>C	2013
HH5	Haplotype HH5	TFB1M	Deletion of 138kbp	2016
HH6	Haplotype HH6	SDE2	A>G	2018
HH7	Haplotype HH7	CENPU	Deletion of 4bp	2020
DUMPS	Deficiency of Uridine Monophosphate Synthase	UMPS	C>T	1993

Table 2. Percentage of carriers of selected genetic diseases in different countries.

Disorder name	N	Healthy	Carriers	% carriers	Country	References
DUMPS	500	500	0	0	Turkey	Korkmaz Agaoglu *et al.*, (2015)
DUMPS	73	73	0	0	Russia	Koshchaev *et al.* (2018)
HH1	248	248	0	0	Brazil	Albertino *et al.* (2022)
HH1	5729	5619	110	1,92	USA	Cole *et al.* (2016)
HH3	14 000	13 286	714	5,1	Germany	Schütz *et al.* (2016)
HH3	17 869	17 869	527	2,95	USA	Cole *et al.* (2016)
HH4	1218	1173	45	0,37	USA	Cole *et al.* (2016)
HH5	2100	1985	115	5,5	Germany	Schütz *et al.* (2016)

Cole *et al.* (2016) at a similar time. Our research results are also in accordance with the results of studies by Schütz *et al.* (2016) (the number of HH3 and HH5 carriers over 5%). Research by Cole *et al.* (2016) shows a lower percentage of HH4 carriers than in Poland in 2019-2020. There are no data available on the number of HH6 and HH7 carriers in other countries. The percentage of carriers of selected genetic diseases in individual countries is presented in Table 2.

Conclusion

In recent years, several causal mutations have been discovered in dairy cattle, most of which negatively affect fertility - causing increased embryo mortality. Most of these diseases are inherited autosomal recessively. From a breeding point of view, it is precisely recessive mutations that have highly negative effects. That kind of diseases occur only in animals that inherit two mutant alleles (homozygotes). Animals with a single altered allele (heterozygote) do not get sick, but can pass the mutations on to their offspring. This makes it much more difficult to diagnose the disease at an early stage and to take preventive measures, e.g. to avoid using these pieces for matings. Therefore it is important to screen the population and identify carries to avoid economic losses due to these genetic disorders in the herd. Thanks to the regular application of genetic tests around the world, the population can be controlled.

Acknowledgement

The authors would like to thank the Breeders for providing the data for the study.

References

Albertino, L. G., Albuquerque, A., Ferreira, J. F., Oliveira, J., Borges, A. S., Patelli, T., and Oliveira-Filho, J. P. 2022. Allele Frequency of APAF1 Mutation in Holstein Cattle in Brazil. Front Vet. Sci. 9: 822224.

Citek, J., Rehout, V., Hajkova, J., Pavkova, J. 2006. Monitoring of the genetic health of cattle in the Czech Republic. Vet. Med. Czech. 51(6): 333–339.

Cole J.B., Null D.J., VanRaden P.M. 2016. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. Journal of Dairy Science. 99(9): 7274-7288.

Debnath, A., Kumar, A., Maan, S., Kumar, V., Joshi, V.G., Trilok N., Sangwan M.L. 2016. Molecular screening of crossbred cow bulls for important genetic disorders. Haryana Vet. 55(1): 93–96.

Korkmaz Agaoglu, O., Agaoglu, A.R., Saatci, M. 2015. Estimating allele frequencies of some hereditary diseases in Holstein cattle reared in Burdur Province, Turkey. Turk. J. Vet. Anim. Sci. 39: 338–342.

Koshchaev, A.G., Shchukina, V., Garkovenko, A.V., Ilnitskaya, E.V., Radchenko, V.V., Bakharev, A.A., Khrabrova, L.A. 2018. Allelic variation of marker genes of hereditary diseases and economically important traits in dairy breeding cattle population. J. Pharm. Sci. Res. 10(6): 1566–1572.

Oner, Y., Keskin, A., Elmaci, C. 2010. Identification of BLAD, DUMPS, Citrullinamia and Factor XI Deficiency in Holstein Cattle in Turkey. Asian J. Anim. Vet. Adv. 5(1): 60–56.

Patel, R.K., Singh, K.M., Soni, K.J., Chauhan, J.B., Sambasiva Rao, K.R.S., 2006. Lack of carriers of citrullinaemia and DUMPS in Indian Holstein cattle. J. App. Genet. 47: 239–242.

Schütz, E, Wehrhahn, C, Wanjek, M, Bortfeld, R, Wemheuer, WE, *et al*. 2016. The Holstein Friesian Lethal Haplotype 5 (HH5) Results from a Complete Deletion of TBF1M and Cholesterol Deficiency (CDH) from an ERV-(LTR) Insertion into the Coding Region of APOB. PLOS ONE 11(6): e0157618.

A time-series analysis of milk productivity changes in US dairy states

M. Li [1] K.F. Reed [2] and V.E. Cabrera[1]

[1]Department of Animal and Dairy Sciences, University of Wisconsin-Madison,
Madison WI 53705, USA
Corresponding Author: vcabrera@wisc.edu
[2]Department of Animal Science, Cornell University,
272 Morrison Hall, Ithaca, NY 14850, USA

Abstract

It is important to characterize trends and seasonal patterns to project amounts and fluctuations in milk and milk components by states or regions. Hence, this study aimed to 1) quantify historical trends and seasonal patterns of milk and milk components production by US states; 2) classify states with similar trends and seasonal patterns into clusters; and 3) summarize the general pattern for each cluster for further applications in simulation models. Our dataset contained 9.18 million lactation records from 5.61 million Holstein cows distributed in 17 states during the period January 2006 to December 2016. Each record included a cow's total milk, fat, and protein yield during a lactation. We used time series decomposition to obtain each state's annual trend and seasonal pattern in milk productivity. Then, we classified states with agglomerative hierarchical clustering into groups according to 2 methods: (1) dynamic time warping on the original time series and (2) Euclidean distance on extracted features of trend and seasonality from the decomposition. Results showed distinguishable trends and seasonality for all states and lactation numbers. The clusters and cluster centroid pattern showed a general upward trend for all yields (ECM, milk, fat, and protein) and a steady trend for fat and protein percent for all states except Texas.

We also found a larger seasonality amplitude for all yields (ECM, milk, fat, and protein) in higher lactation numbers and a similar amplitude for fat and protein percent across lactation numbers. The results could be used for advising management decisions according to farm productivity goals. Furthermore, the trend and seasonality patterns can be used to adjust the production level in a specific state, year, and season for farm simulations to accurately project milk and milk components production.

Keywords: Time series, calving time, production.

Introduction

Dairy cow milk productivity in the US grew by 14%, from 19,895 kg to 22,761 kg, between 2006 and 2016 (USDA-NASS, 2020). Also, there is a recognized seasonal pattern of milk yield and milk components (Salfer *et al*., 2019; Ferreira *et al*., 2020). Moreover, milk yield per cow varies significantly between states and regions (USDA-ERS, 2021). Therefore, it is challenging to predict milk production by location, year, and season. Providing an accurate estimate of lactation performance would help simulation models at the farm and market level that inform management and policy actions.

Milk and milk components productivity according to calving date can be represented as time series. Time series consist of a set of variables and the time-sequential information reliant on them. Time series can be decomposed to identify trends and seasonal patterns (Hyndman and Athanasopoulos, 2018). This study aims to 1) quantify historical trends and seasonal patterns of milk and milk component production in the US; 2) cluster states with similar trends and seasonal patterns; and 3) describe the overall pattern for each cluster for future applications in simulation models.

Material and methods

Dataset

We used a large dataset provided to us by the US Council on Dairy Cattle Breeding (https://www.uscdcb.com/). The dataset included Holstein lactation records with 7 main variables: milk yield (kg/cow for a whole lactation), fat yield (kg/cow for a whole lactation), protein yield (kg/cow per lactation), lactation starting date, parity, US state, and lactation length (d) across an 11-year period (January 2006 to December 2016). We filtered lactation records so that each state had at least 100,000 lactation records, resulting in data from 17 states containing 9,184,086 million lactation records from 5,606,351 million Holsteins. We computed fat percent, protein percent, and ECM (IFCN, 2010) from the milk, fat, and protein yields. Then, we averaged the lactation-length yields for milk, fat, protein, and ECM, as well as the fat percent and protein percent by the week when lactations started.

Time series decomposition

We utilized R (R Core Team, 2020) to decompose the time series of each response using an additive model. The trend component indicated long-term change, the seasonal component reflected a cyclical process, and the residual component reflected short-term influences across the series' period (Hyndman and Athanasopoulos, 2018).

Time series clustering

To categorize all state and parity combinations, we executed agglomerative hierarchical clustering using Ward's minimal variance approach (Murtagh and Legendre, 2014) by minimizing the total variance. We determined the distance between each of the two time series in the original data using dynamic time warping distance and the trend and seasonality components using Euclidean distance. We clustered 18 time series groups into hierarchical cluster trees: 3 types of time series (original, trend, and seasonal) × 6 variables (milk yield, protein yield, fat yield, ECM yield, protein percent, and fat percent). For each one of these 18 time series groups, there were 3 parity groups (1st, 2nd, and 3rd + lactations) ×17 states: 51 time series into one hierarchical cluster tree.

Results and discussion

Each decomposed time series contained yearly trend and seasonality components. Figure 1 shows an example of the original ECM yield for New York state from 3^{rd} + lactations data along with their decomposed components. The clustering method allowed the decomposed variable components to be grouped into clusters based on distance measurements of each pair of two time series. We identified the optimal number of clusters for each group to be 5 by tests on grouping indications. Figure 2 shows the dendrogram of clustering results of ECM yield to illustrate clusters for the original, seasonal, and trend time series.

For the clusters on original data, ECM, milk yield, fat yield, and protein yield had similar patterns for classification where the 1st lactations were mostly classified into 2 clusters that were relatively independent from 2nd and 3rd + lactations. A visual evaluation of the 1st lactation clusters indicated a lower production level and less variability. The 2nd and 3rd + lactations of the same state were commonly classified into the same group, and the centroid patterns of the two major 2nd and 3rd + lactations clusters showed a difference in production levels. The average yield of the 1st lactation and 2nd lactations were 85.4% and 98.3% of the 3rd lactations yields, respectively. The classification pattern for fat percent and protein percent was different from all yields and from each other. Regardless of lactation number, the time series of the same state were clustered together for fat percent and therefore mainly influenced by the geographic factors rather than parity effect. In general, fat percent showed greater variability and less cyclicity than protein percent, suggesting that the protein percent was less sensitive to other factors than the seasonal factors.

For clusters on trends, the classification patterns were like the original data in terms of grouping 2nd and 3rd + lactations from the same state into the same cluster for all yields (ECM, milk, fat, and protein), all lactation from the same state into the same cluster for fat percent, and 1st and 2nd lactations of the same state into the same cluster for protein percent. All yields cluster centroid patterns exhibited a rising. In contrast, the trends for fat and protein percent were relatively steady, and Texas even showed a decline trend in fat percent and protein percent. Florida had the lowest fat percent and protein percent levels, whereas California had the lowest fat percent level, but the highest protein percent for both 1st and 2nd lactations.

For clusters on the seasonal pattern, all the yield clusters had one major cluster for each of 1st, 2nd, and 3rd + lactations. Since our time in each time series was representing the week of calving, we found the valley of the centroid patterns occurred during summer calvings (mainly the last week of July for 2nd lactation, and the 2nd week of July for 3rd lactation), whereas the peak occurred in late January for 2nd lactation and mid-December for 3rd + lactations. For the 1st lactation, milk yields centroid pattern did not vary much before August. All lactations in Virginia, the 2nd lactations in Illinois, and the 2nd and 3rd + lactations in Georgia showed a distinct milk yield centroid pattern in which the yield did not decrease until May. Fat yield in some groups, particularly 3rd +

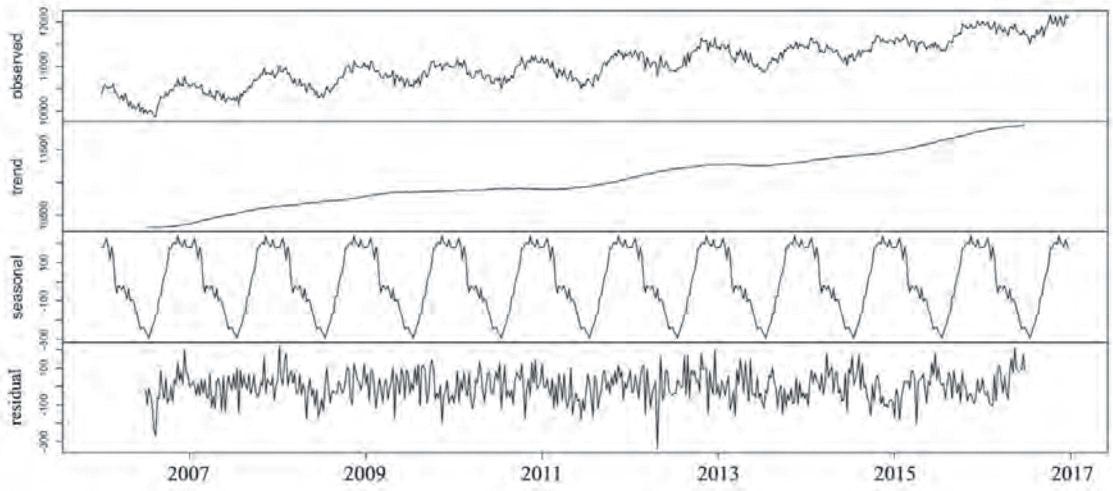


Figure 1. Time series decomposition of ECM yield for 3rd + lactations in New York . All units are in kg/cow/year.

lactations showed a valley prior to late July. Protein yield for 3rd + lactations had a valley between late June and early July. These valleys for components occurred before the milk yield, leading to an earlier valley for ECM. The fat and protein percent, where the 2nd and 3rd + lactations from the same state were clustered together, showed slopes throughout all lactations that did not appear to be different from one another. Peaks of the centroid patterns for fat percent happened during summer weeks (especially early August) and shortly after the protein percent peak. Protein percent exhibited a consistent and clear seasonal pattern across states and lactations, with peaks around late July to early August, a steady decline from November to March or April, and an increase thereafter.

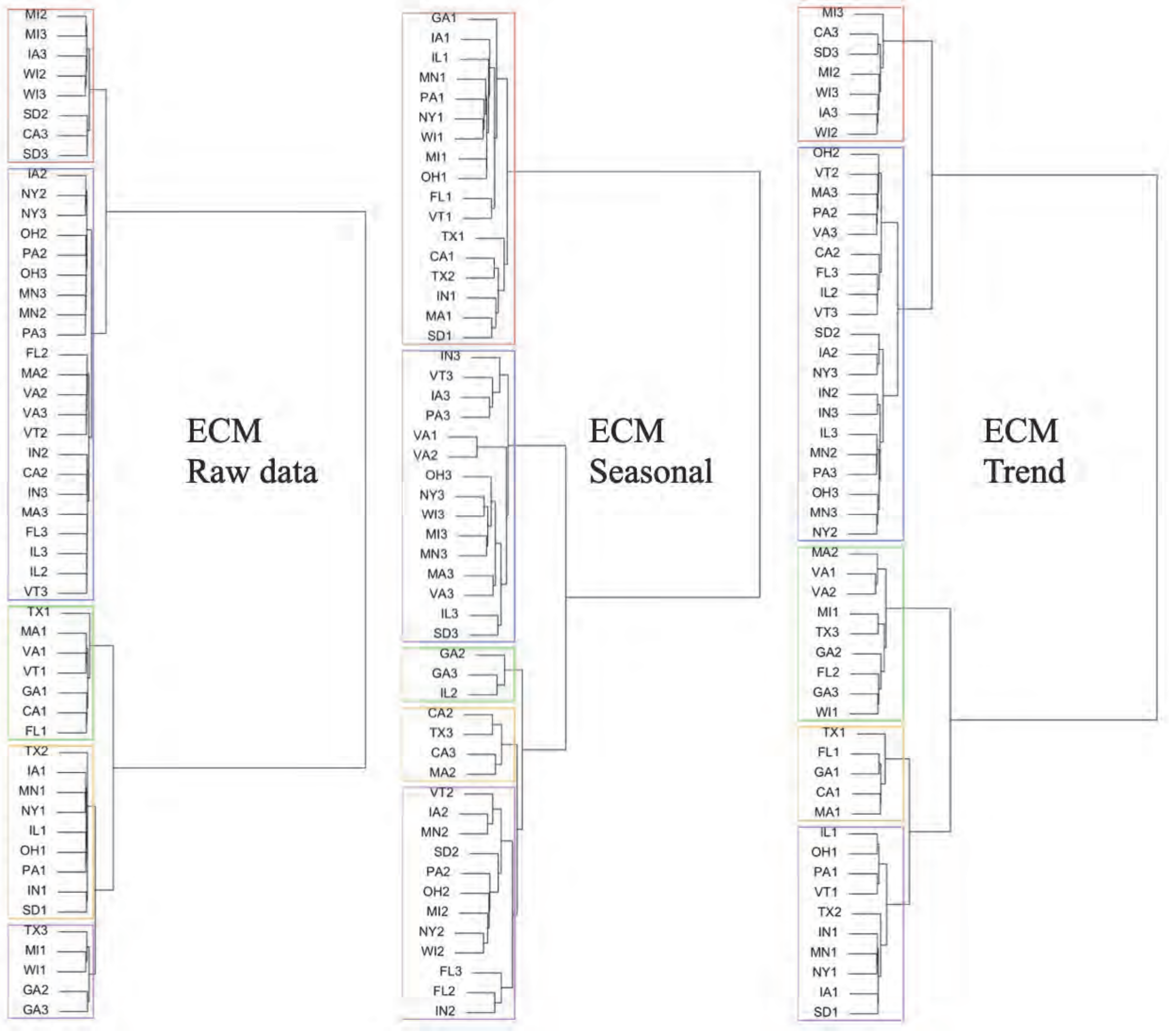


Figure 2. Dendrogram illustrating the groups formed by the cluster analysis for (a) the original data, (b) the seasonal, and (c) the trend of time series of ECM on calving date for each state and parity group. MI2 = 2nd lactation in Michigan.

Conclusion

Time series of all yields (ECM, milk, fat, and protein) were largely influenced by the long-term effect of the trend and the periodic effect of seasonality. Analysis of the long-term trend showed a general upward pattern across the years for all yields and a general flat pattern across the years for fat and protein percent. Analysis of seasonal trends showed a stronger effect of seasonality for later lactations for all yields and that seasonality of fat and protein percent are not influenced by parity. Cluster results showed distinct groups with closer long-term trend and seasonality patterns among lactations and locations. The improved classification of the original and the decomposition parameters can assist with herd and cow management decisions demonstrating the importance of seasonal patterns in production variables according to geographical location and parity.

References

Hyndman, R. J., and Athanasopoulos, G. 2018. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2

IFCN Dairy Research Center. 2010. IFCN Dairy Report 2010 – For a Better Understanding of Milk Production World-Wide. 214 pages, IFCN – International Farm Comparison Network, Kiel

Ferreira, F. C., J. S. Clay, and A. De Vries. 2020. Distribution of seasonality of calving patterns and milk production in dairy herds across the United States. J. Dairy Sci 103:8161–8173.

Murtagh, F. and P. Legendre. 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J. Classif. 31(3): 274-295.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Salfer, I. J., C. D. Dechow, and K. J. Harvatine. 2019. Annual rhythms of milk and milk fat and protein production in dairy cattle in the United States. J. Dairy Sci. 102:742–753.

USDA-ERS. 2021. Milk cows and production by state and region (Annual). https://www.ers.usda.gov/data-products/dairy-data/

USDA-NASS. 2020. Agricultural Marketing Service, and Economic Research Service calculations. Accessed May 29, 2020. https://downloads.usda.library.cornell.edu/usda-esmis/files/c821gj76b/jm2159057/qv33sh51q/agpr0520.pdf

Predicting dairy herd resilience on farms with conventional milking systems

R.S.C. Rikkers[1], B. Ducro[2] and C. Kamphuis[1]

[1]Wageningen Livestock Reseacrh, Wageningen University and Research,
P.O. Box 338, 6700 AH Wageningen, The Netherlands
Corresponding Author: roxann1.rikkers@wur.nl
[2]Animal Breeding and Genomics, Wageningen University and Research,
P.O. Box 338, 6700 AH Wageningen, The Netherlands

Abstract

Individual cow resilience depends on the capacity of cows to respond to environmental disturbances. Together with management decisions, that affect the performance of these cows, and their environment they represent herd resilience. Until now, herd resilience can be estimated with the use of daily milk yield observations from an automated milking system (AMS), leaving conventional milking farms (CMS) without information on herd resilience. Therefore, this study investigated the possibility to predict dairy herd resilience using herd performance data generally available on AMS and CMS farms. Data from 585 Dutch AMS farms including herd performance variables: herd size, kg milk, proportion acidosis, proportion ketosis were used to predict herd resilience. As prediction model, a 5- fold cross validation Random Forest model with an extensive grid to finetune model parameter settings was used. Results show that, on average, herd resilience is predictable. Both the mean herd resilience as well as the predicted mean herd resilience were 1.30. The range, however, was much wider for herd resilience (0.70-1.86) than for the predicted herd resilience (1.10-1.50). Pearson correlation between herd resilience and predicted herd resilience was 0.55 ± 0.06. Thus, using only herd performance data that are generally available on farms (AMS and CMS), it is possible to predict herd resilience but not with a high accuracy.

Keywords: Dairy herd resilience, random forest, prediction, automated milking system.

Introduction

Individual cow resilience indicates that cows are minimally affected by environmental disturbances, such as pathogens or extreme weather, and their production quickly recovers if they are affected (Colditz and Hine, 2016). The individual cow resilience together with management decisions that affect the performance of these animals and their environment represents herd resilience (Blanc *et al.*, 2013). For dairy herds, this could be defined as that resilient herds show less milk yield deviations on herd level and thereby the herd as a whole is assumed to be less affected by disturbances. Individual cow resilience (Elgersma *et al.*, 2018; Poppe *et al.*, 2020) and dairy herd resilience can only be estimated on farms that use automated milking systems (AMS). For individual cow resilience, an expected lactation curve is fitted and less deviations from this fitted lactation curve indicates good individual resilience (Poppe *et al.*, 2020). The estimated individual resilience is corrected for a fixed herd-year and year-season effect and a random genetic and error effect. This fixed herd-year effect is assumed to represent herd resilience and it has been shown to correlate with herd management decisions (Poppe *et al.*, 2021). Estimation of individual or dairy herd resilience is only possible with the use of daily milk yield observations from AMS, and only after a full lactation. Still, a large portion of farms do not have an AMS but milk conventionally (in a milking parlour). Therefore, this study investigates the possibility to predict herd resilience with the use of a Random Forest model and herd performance variables usually available on farms with and without AMS.

Material and methods

An existing dataset including 2,644 Dutch AMS farms between 2011 and 2017 was available and consisted of a herd resilience indicator (fixed herd-year effect) and variables describing herd performance: kg milk, fat percentage, protein percentage, proportion with elevated somatic cell count, proportion ketosis, proportion of survival till second lactation, parity, herd size, etc. Incomplete records were removed resulting in a subset of 585 herds between 2012 and 2016 with complete information. These five herd-year estimates were averaged per herd, resulting in a mean herd resilience indicator per herd, which was assumed to represent herd resilience of a farm between 2012 and 2016. Mean herd resilience was 1.30 ranging from 0.70 till 1.86, where low values indicates good herd resilience.

To predict herd resilience, a Random Forest model based on the algorithm of Breiman (2001) was used (R package RandomForest (Liaw and Wiener, 2002). A Random Forest model is a combination of randomly generated decision trees. Each individual tree performs quite poorly, but many trees combined into a ‘forest’ provides more reliable prediction results. Previously, the Random Forest algorithm has shown to produce reliable predictions with e.g. the prediction of dairy cow survival till second lactation (Heide *et al,*. 2019) and with the prediction of lifetime resilience in dairy cows (Ouweltjes *et al.*, 2021). Finetuning of the Random Forest model parameters was done using an extensive grid: number of generated trees were 500, 1000, 1500, 2000 and 5000, minimum number of herds per branch were 5, 10 and 15, and the random number of candidate variables at each split 1 till 34 (in total 34 predictive variables were available). Each of the 510 model combinations was trained on 80% of the data and validated on the remaining 20% of the data using a 5-fold cross validation stratified to herd. This means that each herd was used four times in the training data and once in the validation set.

Results and discussion

Random Forest model finetuning

The best performing Random Forest model combination, based on the Pearson correlation between herd resilience and predicted herd resilience (0.55 ± 0.06), was the model that generated 500 random trees, contained a minimum number of five herd per branch and that randomly selected four predictive variables per split. The poorest performing model combination was the model that generated 1500 trees, contained a minimum number of 15 herds per branch and randomly selected one predictive variable per split (Pearson correlation estimated vs. predicted 0.52 ± 0.06). Considering that 500 randomly generated trees and a minimum of five herds per branch are default parameters of the randomForest function (Liaw and Wiener, 2002) and the difference between the best and poorest performing model is small, the need for an extensive grid was not necessary for this dataset.

Table 1. Mean, minimum and maximum of estimated and predicted herd-year estimates.

Herd-year effect	Mean	Min – Max
Estimated	1.30	0.70 – 1.86
Predicted	1.30	1.10 – 1.50

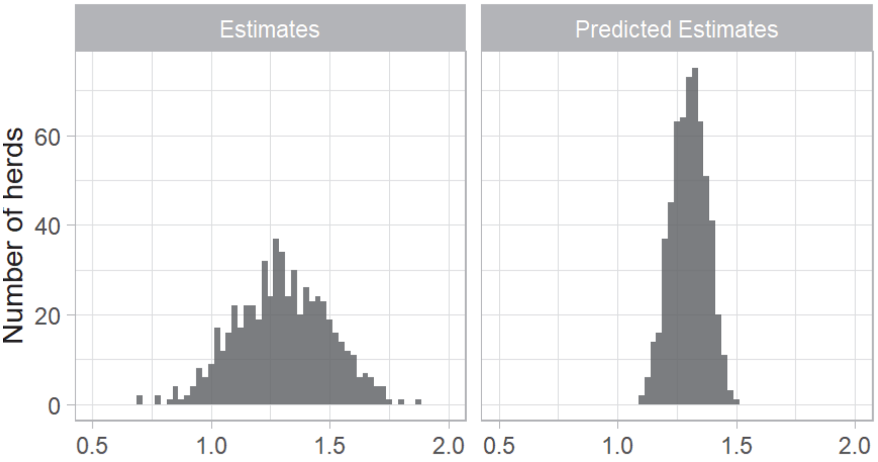


Figure 1. Histogram of herd-year estimates (range 0.70-1.86) and predicted herd-year estimates (range 1.10-1.50) using a Random Forest model.

Prediction accuracy

Pearson correlation between herd resilience and predicted herd resilience was 0.98 ± 0.00 for the 80% training datasets and 0.55 ± 0.06 for the 20% validation datasets (mean and standard deviation of the five k-folds). On average the Random Forest model performed quite well, since the average predicted herd resilience indicator of 1.30 was similar to the estimated herd resilience indicator (Table 1). However, the range of herd resilience was 0.70-1.86 while the range of predicted herd resilience was 1.10-1.50 (Table 1; Figure 1).

A method to improve prediction accuracy could be discretization, which is a pre-processing step where continuous variables are transformed to discrete variables. Either transforming the continuous predictive variables or the continuous variable herd resilience might improve the prediction accuracy, since applying discretization in complex datasets has been shown to significantly improve model performance (Lustgarten *et al.*, 2008).

Conclusion

Using only herd performance data that is generally available on farms it is possible to predict herd resilience but not with a high accuracy.

Acknowledgement

We acknowledge European Union's Horizon 2020 research and innovation program (GenTORE) under grant agreement No. 727213 for their financial support.

References

Blanc, F., E. Ollion,, L. Puillet, L. Delaby, S. Ingrand, M. Tichit, and N.C. Friggens, 2013. Evaluation quantitative de la robustesse des animaux et du troupeau: quels principes retenir. Proceedings of the 20th Rencontres Autour des Recherches sur les Ruminants, 4-5.

Breiman,, L., 2001. Random Forests. Machine Learning 45, 5–32

Colditz, I. G., and B.C. Hine, 2016. Resilience in farm animals: Biology, management, breeding and implications for Animal Welfare. Anim. Prod. Sci. 56(12), 1961.

Elgersma, G. G., G. de Jong, R. van der Linde and H.A. Mulder, 2018. Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. J. Dairy Sci. 101(2), 1240–1250.

Heide, E.M.M. van der, R.F. Veerkamp, M.L. van Pelt, C. Kamphuis, I. Athanasiadis and B. Durco, 2019. Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. J. Dairy Sci. 102:9409-9421.

Liaw A. and M. Wiener, 2002. Classification and Regression by randomForest. R News 2(3), 18—22.

Lustgarten, J.L., V. Gopalaksirhnan, K. Grover and S. Viswerswaran, 2008. Improving classification preformance with discretization on biomedical datasets. AMIA Annu Symp. Proc. 2008 445-449

Ouweltjes W., M. Spoelstra, B. Ducro, Y. de Haas and C. Kamphuis, 2021. A data-drive prediction of lifetime resilience of dairy cows using commercial sensor data collected during first lactation. J. Dairy Sci. 104:11759-11769.

Poppe, M., R.F. Veerkamp, M.L. van Pelt and H.A. Mulder, 2020. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. J. Dairy Sci., 103(2), 1667–1684.

Poppe, M., H.A. Mulder, C. Kamphuis and R.F. Veerkamp, 2021. Between-herd variation in resilience and relations to Herd Performance. J. Dairy Sci., 104(1), 616–627.

Imputation of missing test day milk records and its use in genetic evaluation for milk yield in Mehsana buffaloes

Sujit Saha, Swapnil Gajjar, Nilesh Nayee, A. Sudhakar, G. Kishore and R.O. Gupta

National Dairy Development Board, Anand-388001, Gujarat, India

Abstract

Under the smallholder dairying, the establishment of robust performance recording infrastructure for genetic evaluation programmes is quite challenging. The nationwide lockdown due to COVID-19 pandemic hampered the field performance recording activities to a great extent.The restriction in mobility of the field functionaries leads to the accumulation of considerable volume of missing TestDay Records (TDR). As the number of TDR is important for reliable breeding value (BV) estimates, the accumulation of missing values under any genetic evaluation program would be a concern for the implementers. Under such circumstances, imputation of TDR could provide an effective solution.The present study was carried out to assess the accuracy of the imputation of the missing TDR using Linear and Cubic spline interpolation in Mehsana Buffaloes and the effect of using this imputed TDR for BV estimation.The results of this study indicated a high correlation (0.89) between actual and imputed TDR. The BV estimates and their reliabilities obtained using a combination of actual and imputed TDR were found to be at par with that estimated using 100 per cent actual TDR.There was no impact on the ranking of Mehsana bulls. The study revealed that imputation of missing record through interpolation holds the potential of alternate performance recording system with bi-monthly/quarterly interval without compromising with the accuracy. Besides, it could also facilitate to cover more animals under recording with the limited available fund and create a reference population with a wide genetic base for successful implementation of genomic selection using recorded females.

Keywords: Mehsana buffaloes, missing values, interpolation, breeding values.

Introduction

Low capital investment, short operating cycle, steady returns etc. made dairying a preferred supplementary livelihood option for rural households in India. It has been contributing to the farmers in many ways like regular income from milk and milk products, insurance against drought, emergency cash requirements, household nutrition, fuel for cooking, manure for crops, draught power for farming etc. (Rath, 2019). In India, ownership of bovines is fragmented, with a large number of small and marginal farmers, each raising a few animals (with the average herd size of 5 or below) for draught (animal traction) or dairy purposes. Hence, implementation of long term and scientific genetic evaluation programmes through systematic performance recording is a quite challenging and costly proposition. It is estimated that, to implement a progeny testing programme, based on a young sire model, involving small herds of the dairy farmers, around $ 0.57 million (INR 40 million) per annum would be required to carry out Test Artificial Insemination of about 20 young bulls and other related field activities like Nominated AI, milk recording, measuring type traits, growth monitoring of daughters at six-monthly intervals and also for overall monitoring and supervision activities. It is also estimated that out of the total budget, around 30-40% of the total fund is being utilized for the test day milk recording and milk component analysis at a monthly interval.

The COVID-19 pandemic, which has emerged as one of the biggest pandemics in and around the globe has a devastating effect not only on health but also the global economy (Pathak, 2020). Due to nationwide lockdown, like in many industries, many routine activities under ongoing field-based genetic improvement programs were also got affected. To prevent the spread of infection, farmers even refused to allow AI technicians or milk recorders to enter their premises. Thus activities like AI delivery, monthly milk recording, milk sampling etc. were either stopped or carried out at the bare minimum level. It resulted into a considerable amount of data loss, till the activities resumed at its normal pace. Similar situation arises even during other unavoidable circumstances like natural calamities.

The possible adverse impact of missing records in Breeding value (BV) estimates of the animals and it's reliability, triggered a thought process to adopt a suitable alternative, which enables estimation of breeding values of the animals having less number of test day (TD) records without compromising the reliability of estimates.

With this background, the present study was designed to assess the efficiency of different interpolation approaches to impute missing TD records (TDR) and impact of using such imputed records in estimating BV for milk yield.

Materials and methods

About the "Mehsana" buffaloes

Mehsana buffaloes are one of the best milk breeds of buffalo in India (Gupta, 1997) and are spread in the northern part of the Gujarat State mainly Mehsana, Banaskantha and Sabarkantha districts (Prajapati *et al.*, 2018).

The name of Mehsana buffalo was derived from the town "Mehsana" in the North Gujarat State. The Mehsana Buffalo breed was evolved by crossing Surti and Murrah buffaloes by the farmers to meet the need for higher milk and also for adaptation to the semi arid climate of Gujarat (Pundir *et al.*, 2000).

Mehsana buffalo is recognized as a persistent milker, regular breeder and most economical to adverse climatic condition. According to the National Bureau of Animal Genetic Resources (NBAGR), the average lactation yield of Mehsana buffaloes is about 1988 kg, with maximum yield reported to be 3597 kg (www.nbagr.res.in).

Brief about genetic improvement program on Mehsana buffaloes

For Genetic improvement of Mehsana Buffaloes, NDDB initiated its first field-level progeny testing project (PT) in the year 1987 in association with The Mehsana District Cooperative Milk Producers' Union Limited, Mehsana, Gujarat under Dairy Herd Improvement Programme Actions (DIPA). Under this project, a robust infrastructure was created for milk recording and genetic evaluation (Trivedi, 1997). Trained milk recorders were engaged to record test day milk yield at the monthly interval and collect milk samples for analysis of various milk components till 2012, the project tested about 231 Mehsana bulls. Since 2012-13 to 2018-19, the PT activities were carried out under the supervision of NDDB in National Dairy Plan Phase–I, a World Bank funded central sector scheme of Govt. of India.

Source of data

The present study was carried out using 119066 test day milk records obtained from 11962 Mehsana buffaloes recorded during September 1988 to September 2019 under Meshana Buffalo PT project implemented by The Mehsana Milk Cooperative Union Limited, Mehsana, Gujarat. The TDR for milk were retrieved from NDDB's INAPH database. For the current study, only those Mehsana buffaloes having 3 or more test day records have been considered. Besides this, observations (with respect to various traits) falling outside the defined physiological range (outliers) were removed. Criteria for accepting records for analysis is given in Table 1.

Imputation of missing records through interpolation

Under this study to mimic the actual field scenarios pertaining to the missing test day records, actual test day records of the buffaloes were masked following four different approaches namely ST-1, ST-2, ST-3, ST-4 as shown in Figure 1. The number of records masked under different strategies is elaborated in Table 2.

The individual animal-wise masked records were imputed using two different interpolation methods namely Linear interpolation and Cubic spline using zoo package (Zeileis *et.al*, 2020) of R software. The degree of association between actual test-day records and imputed test day records was assessed through correlation estimate. The imputation accuracy of linear interpolation and cubic spline interpolation was expressed in terms of Mean Absolute Error (MAE) and Root Mean Sum Square Error (RMSE).

Breeding value estimation

BV along with it's reliability of the recorded she buffaloes and buffalo bulls were estimated using DMU software (Madsen and Jensen, 2014) applying random regression (with legendry polynomials) test day model as mentioned below:

$$y_{thijmnkl} = Village_h + YS_i + Age_j + HYMR_m + Owner_n + \sum_{l=0}^{nf} \emptyset_{ktl}\beta_l + \sum_{l=0}^{nr} \emptyset_{ktl}u_{kl} + \sum_{l=0}^{nr} \emptyset_{ktl}pe_{kl} + e_{thijmnkl}$$

where,

$y_{thijmnkl}$ is the test-day milk yield of animal *k* recorded on day *t* within fixed village subclass *h*, fixed YS (year of calving x season of calving) subclass *i,* fixed Age at calving (6 months grouped) subclass *j,* random HYMR (herd x year of recording x month of recording) subclass m and random owner (at the time of first milk recording) subclass n;

β_l are fixed regression coefficients;

u_{kl} and pe_{kl} are the l^{th} random regression for animal additive genetic and permanent environmental effects, respectively, for animal *k;*

Φ_{ktl} is the l^{th} legendre polynomial for the test day record of cow *k* made on t^{th} day in milk;

nf is the order of polynomials fitted as fixed regressions (0 to 2);

nr is the order of polynomials for *u* and *pe* effects (0 to 2);

$e_{thijmnkl}$ is the random residual effect.

Table 1. Acceptable Physiological range for accepting records for analysis.

SN	Traits	Acceptable Range
1.	Days in Milk	5-330 days
2.	Test Day milk yield	1-40 kg
3.	Age at first calving	20-80 months

Table 2. Strategies for Masking of Test day records for different groups of animal.

Strategy	Animals considered	No. of eligible buffaloes in the subset	Total TD records	No. of TD records masked & subsequently imputed	Remarks
ST-1	Having 10 TD records	2915	29150	11660	Alternate TD records masked (40% of the total TD records)
ST-2	Having 10 TD records	2915	29150	17490	alternate two TD records masked (60% of the total TD records)
ST-3	Having 3 to 10 TD records	6120	54495	29545	54% of the total TD records masked
ST-4	Having 3 to 11 TD records	11916	118252	64321	54% of the total TD records masked

Table 3. Correlation coefficient between true and imputed test day milk records.

	Imputation method	
Strategy	Linear Interpolation	Cubic spline
ST-1	0.899*	0.897*
ST-2	0.890*	0.888*
ST-3	0.885*	0.867*
ST-4	0.883*	0.876*

For analysis of each subset of animals, two types of data files were constructed. One with 100% actual test day record TDR and another with the combination of actual and imputed TDR.

The degree of association of BVs (for both the she buffaloes and breeding bulls, respectively) predicted using the combination of imputed and actual TDR with the BV estimates obtained using 100% actual TDR was studied through Pearson's correlation coefficient (Snedecor and Cochran, 1989). In addition to that ranking of these animals based on BV was also compared using Spearman's rank correlation coefficient (Steel and Torrie, 1960).

The schematic diagram of the design of the experiment is elaborated in Figure 2.

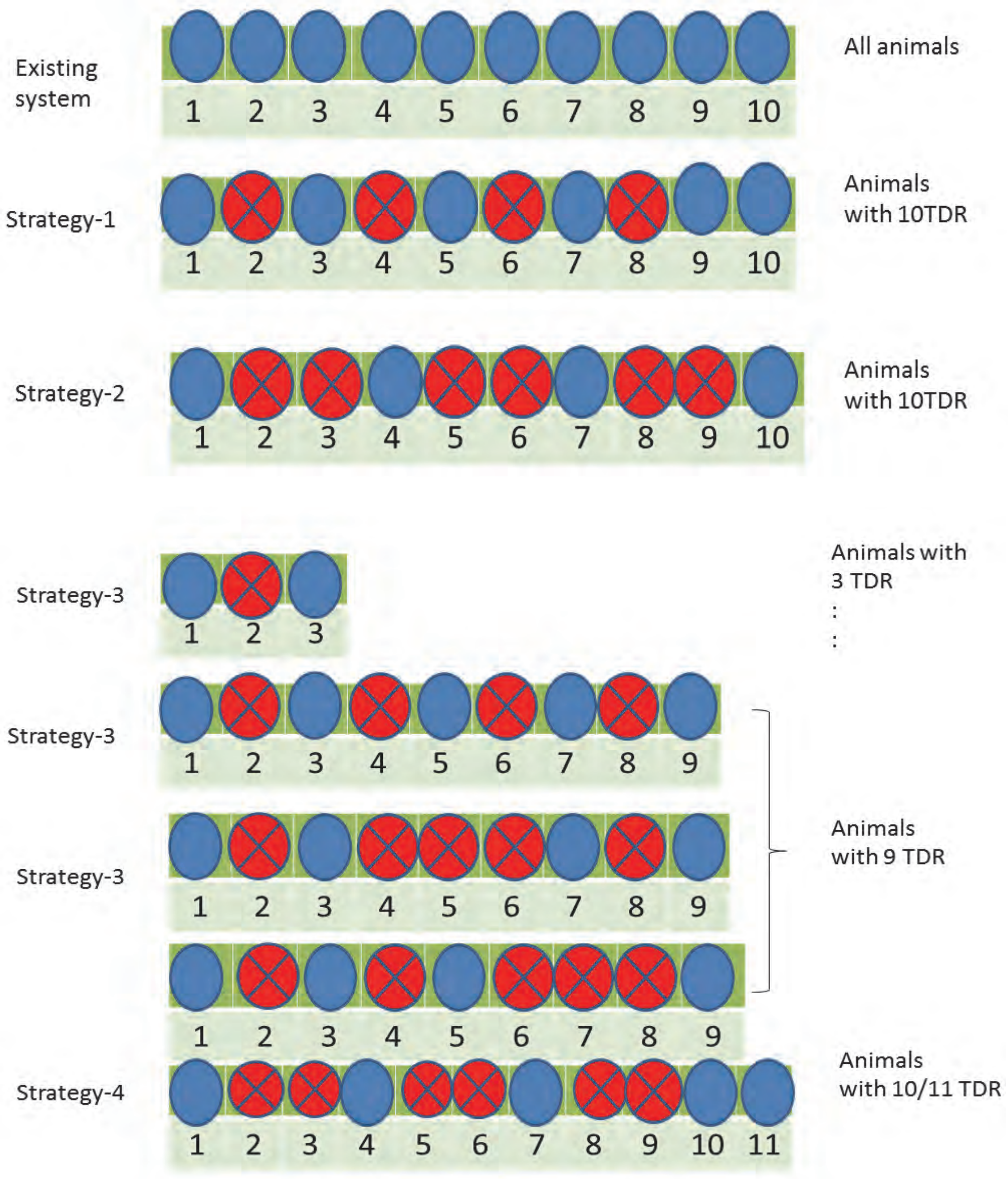


Figure1. Test Day Milk record (TDR) masking strategies. Similarly animals having TDR between 4-8 are masked in different possible combinations.

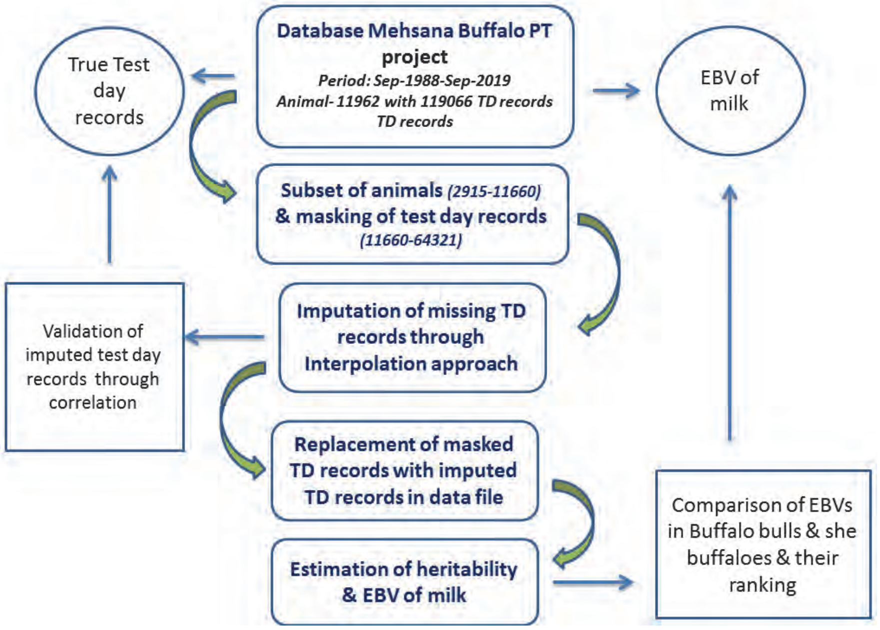


Figure 2. Design of Experiment- a schematic diagram.

Results and discussion

The correlation between imputed TDR (obtained using Liner interpolation and cubic spline method) and the actual test day records are presented to Table 3.

The obtained result indicated a statistically significant and higher degree of association (correlation coefficient >0.86) between imputed TDR with that of actual TDR in Mehsana buffaloes. However, the correlation was observed to decline when the animals having <10 TDR (i.e. 3 to 9 TD records) were masked for imputation.

Comparison of the Linear and Cubic spline interpolation approach expressed in terms of the correlation between actual TDR and imputed TDR and it was found to be marginally higher in linear interpolation than cubic spline interpolation.

The imputation accuracy of Linear interpolation and Cubic spline interpolation expressed in terms of MAE and RMSE was found to be very close (Table 4). However, the MAE and RMSE values were found to be marginally lesser in case of Cubic spline approach.

The BV of the recorded she buffaloes, as well as breeding bulls, was estimated using actual TDR and the combination of imputed and actual TDR, separately. Heritability estimates of test day milk yield obtained under different masking strategies were given below in Table 5.

As presented in Table 5, the heritability estimate of milk yield using 100% actual TDR was found to be 0.26. However, with the use of a combination of imputed missing data along with actual test day records, heritability estimates were observed to reduce marginally (0.24-0.25), which may be due to loss of variance in the test day records due to imputation using interpolation. Heritability estimate thus obtained in this study was found to be slightly higher than the estimate reported by Galsar *et al.* (2016) and Prajapati *et al.* (2018) in Mehsana Buffaloes, whereas, Breda *et al*. (2010) reported TD milk yield heritability between 0.19 -0.31 in Murrah buffaloes.

The correlation coefficient between the BV estimates obtained using 100% actual TDR and a combination of true test day cum imputed TDR are presented in Table 6 (for she buffaloes) and Table 7 (for breeding bulls), respectively.

The results indicated that the replacement of a part of actual TDR with imputed TDR did not have any adverse impact on BV estimation. In case of she buffaloes, the correlation was observed to vary between 0.97 to 0.99 under various strategies for both linear interpolated as well as cube spline interpolated data.

While in the case of breeding bulls, the correlation coefficient for all the cases were observed to be 0.99. All the correlation coefficient estimates were found to be statistically significant. The correlation between the rankings of the animals as estimated by spearman's rank correlation were found to be very high (around 0.98).

Ranking of the top 10 Mehsana breeding bulls based on BV calculated using actual TDR and combination of actual and imputed TDR are presented in Table 8 and Table 9, respectively for linear interpolation as well as spline cube interpolation. The results revealed that there was no marked change in the BV estimates as well as the ranking of the breeding bulls due to incorporation of imputed TDR for data analysis.

Panchal *et al*. (2019), studied the impact of different milk recording strategies on sire evaluation in HF crossbred cattle and found monthly test day milk recording as an optimum strategy for milk recording in genetic evaluation program and concluded that a bimonthly recording can be considered with caution under the conditions wherein deploying monthly test day milk recording is not feasible. The rank correlation between sire breeding values for 305 days milk yield was reported to be 0.79 between monthly milk recording and bimonthly recording strategies, which was lower than the estimates observed in the current study (>0.97). It indicated that in case of presence of the missing TDR, imputation of the missing values with interpolation and its use in breeding value estimation was quite effective in ensuring greater reliability of estimate rather than excluding the same from the analysis.

Conclusion

Based on the results obtained in the present study, it can be concluded that under adverse situations like COVID-19 pandemic or other natural calamities, when regular monthly milk recording activities in the field level are not feasible, imputation of the missing records can be carried out through a suitable interpolation approach, and such imputed data can be effectively utilized in combination of available actual records for genetic evaluation of animals without any adverse impact on breeding value estimation and its reliability.

Under smallholders' dairy production systems, where dairy animals of important breeds are spread across the country under varied environmental conditions, covering a maximum number of animals under performance recording is quite challenging and a costly affair. In such conditions, a performance recording system with the frequency of recording at bi-monthly or quarterly interval holds the potential to include of maximum possible number of animals of a particular breed under milk recording program with the available funds and also to study the influence of Genetic x Environment interaction, In such scenarios, the missing monthly records could be imputed through interpolation and used for genetic evaluation purpose. Further, bringing the maximum number of animals under systematic milk recording program would also enable the creation of a large reference population with a wide genetic base for successful implementation of Genomic selection using recorded females.

Table 4. MAE and RMSE estimate for Linear and Cubic Spline Interpolation.

Interpolation approach	MAE	RMSE
Linear Interpolation	0.713	0.996
Cubic Spline	0.706	0.995

Table 5. Heritability estimates (h^2) of test day milk yield with 100 per cent actual test day record vis a vis combination of actual & imputed test day records.

		h^2	
Strategy	**Type of TDR used**	**Linear Interpolation**	**Cubic spline**
ST-1	Actual + Imputed	0.25	0.25
ST-2		0.25	0.24
ST-3		0.24	0.24
ST-4		0.26	0.25
All animals	100% actual	0.26	

Table 6. Correlation coefficient between EBVs obtained using actual and imputed Test day records in Mehsana buffaloes.

	Pearson correlation		Rank correlation	
Strategy	**Linear Interpolation**	**Cubic spline**	**Linear Interpolation**	**Cubic spline**
ST-1	0.997*	0.997*	0.996*	0.996*
ST-2	0.993*	0.994*	0.992*	0.993*
ST-3	0.977*	0.971*	0.973*	0.968*
ST-4	0.980*	0.982*	0.977*	0.979*

(*P<0.05)

Table 7. Correlation coefficient between EBVs obtained using actual and imputed Test day records in Mehsana Breeding Bulls.

	Pearson correlation		Rank correlation	
Strategy	**Linear Interpolation**	**Cubic spline**	**Linear Interpolation**	**Cubic spline**
ST-1	0.999*	0.999*	0.998*	0.999*
ST-2	0.996*	0.997*	0.995*	0.996*
ST-3	0.991*	0.991*	0.989*	0.990*
ST-4	0.982*	0.986*	0.977*	0.982*

(*P<0.05)

Table 8. Breeding value estimates of top ten ranked Mehsana breeding bulls under different strategies using Linear interpolation of missing records.

	Based on Actual Test day records			ST-1			ST-2			ST-3			ST-4		
Bullid	**BV**	**Accuracy**	**Rank**	**BV**	**Accuracy**	**Rank**	**BV**	**Accuracy**	**Rank**	**BV**	**Accuracy**	**Rank**	**BV**	**Accuracy**	**Rank**
Bull-1	467.84	78.03	1	465.18	78.27	1	462.36	78.54	1	472.32	78.55	1	466.83	79.38	1
Bull-2	416.12	85.93	2	406.14	86.15	2	393.36	86.40	2	386.74	86.48	2	389.59	86.98	2
Bull-3	395.71	84.49	3	380.31	84.60	3	374.50	84.73	4	371.52	84.64	4	365.67	85.21	3
Bull-4	383.78	79.52	4	376.85	79.77	4	376.98	80.06	3	373.26	80.11	3	362.71	80.87	4
Bull-5	367.67	88.04	5	371.16	88.13	5	366.32	88.25	5	369.40	88.18	5	348.29	88.62	6
Bull-6	342.92	54.34	6	338.81	55.80	7	329.18	57.22	9	291.56	58.70	10	290.86	60.32	11
Bull-7	342.92	88.02	7	341.52	88.15	6	332.85	88.30	8	332.48	88.30	7	315.34	88.73	9
Bull-8	341.86	81.37	8	334.42	81.68	9	341.26	82.01	6	334.39	82.17	6	353.13	82.85	5
Bull-9	339.45	88.81	9	336.81	88.86	8	341.05	88.93	7	331.31	88.82	8	324.22	89.22	8
Bull-10	333.67	79.68	10	318.14	79.89	10	320.67	80.15	10	324.31	80.16	9	346.26	80.92	7

Table 9. Breeding value estimates of top ten ranked Mehsana breeding bulls under different strategies using Cubic spline interpolation of missing records.

	Based on actual test day records			ST-1			ST-2			ST-3			ST-4		
Bullid	**BV**	**Accurcy.**	**Rank**	**BV**	**Accurcy.**	**Rank**	**BV**	**Accurcy.**	**Rank**	**BV**	**Accurcy.**	**Rank**	**BV**	**Accurcy.**	**Rank**
Bull-1	467.84	78.03	1	463.44	78.21	1	459.88	78.35	1	473.15	78.45	1	479.57	78.36	1
Bull-2	416.12	85.93	2	404.97	86.14	2	392.31	86.30	2	397.17	86.43	2	404.78	86.37	2
Bull-3	395.71	84.49	3	381.06	84.56	3	383.58	84.61	3	376.56	84.61	4	371.73	84.56	3
Bull-4	383.78	79.52	4	375.45	79.73	4	375.77	79.88	4	380.65	80.01	3	369.51	79.93	4
Bull-5	367.67	88.04	5	368.79	88.11	5	362.76	88.17	5	362.78	88.19	5	345.30	88.14	6
Bull-6	342.92	54.34	6	338.26	55.68	8	329.23	56.76	9	287.20	57.95	11	281.13	57.70	14
Bull-7	342.92	88.02	7	340.40	88.13	6	334.37	88.22	8	327.48	88.27	8	311.56	88.22	9
Bull-8	341.86	81.37	8	332.19	81.64	9	338.45	81.86	7	326.23	82.06	9	357.60	81.97	5
Bull-9	339.45	88.81	9	338.33	88.84	7	347.43	88.85	6	340.50	88.82	6	334.10	88.79	8
Bull-10	333.67	79.68	10	319.27	79.85	10	327.42	79.98	10	328.80	80.08	7	344.97	80.00	7

Acknowledgement

The authors acknowledge the contribution of Mehsana PT project officials and Field staff for their sincere efforts in capturing performance data of Mehsala buffaloes in the field.

References

Breda F. C., Albuquerque L. G., Euclydes R. F., Bignardi A. B., Baldi F., Torres R. A., Barbosa L., and Tonhati H. 2010. Estimation of genetic parameters for milk yield in Murrah buffaloes by Bayesian inference . Journal of Dairy Science. TBC.1-8. doi:10.3168/jds.2009-2230

Galsar N. S. , Shah R. R., Gupta J. P. , Pandey D. P. , Prajapati K. B. and Patel J. B. 2016. Analysis of first production and reproduction traits of Mehsana buffaloes maintained at tropical and semi-arid region of Gujarat, India. Life Sciences Leaflets, 77, 65 to 75. Retrieved from http://petsd.org/ojs/index.php/lifesciencesleaflets/article/view/1076

Gupta P. R. (Ed.)1997. Dairy India, Fifth Edition, Statistics, 153-195.

Madsen P., Jensen J., Labouriau R., Cristensen O.F. and Sahana G. 2014. DMU- a package for analyzing multivariate mixed models in quantitative genetics and genomics. In: Proc. 10th World Congress of Genetics Applied to Livestock Produciton, Vancouver, Canada.

Rath D. 2019. Keynote address at 47th Dairy Industry Conference held at Patna during Feb 7-9, 2019: Indian Dairyman.73(3):16-21.

Panchal D., Kakati P., Joshi R. S., Patel A. C., Dholariya S. A., Shah D. B., Patel S. B., Patel C.T., Gajjar S. G., Kishore G. and Rank D. N. 2019.Effect of Alternative Milk Recording Strategies on Genetic Evaluation of Sires of Holstein Friesian Crossbred Cattle. International Journal of Livestock Research.9(8):203-213.

Pathak K. M. L. 2020. Impact of COVID-19 on health of dairy animals. Indian Dairyman.72(5):50-52.

Prajapati B M, Gupta J P, Chaudhari J D, Parmar G A, Panchasara H H, Chauhan H D, Ankuya K J and Prajapati M N. 2018. First lactation production performance of Mehsana buffaloes under field progeny testing programme in semi-arid region of Gujarat. Indian J of Dairy Science. 71(4): 404-408.

Pundir R., Sahana G., Navani N., Jain P., Singh D., Kumar S. and Dave A. 2000. Characterization of Mehsana Buffaloes in India. Animal Genetic Resources Information. 28: 53-62. Doi:10.1017/S101423390000136X.

Steel R.G.D andTorrie J.H. 1960.Principles and Procedure of Statistics with Special Reference to the Biological Sciences. McGrew Hill Book Company Inc. New York. 550

Trivedi K.R. 1997. A case study of Buffalo recording systems under the Dairy Cooperative organizations in India. International Workshop on Animal Recording for smallholders in developing countries. Anand, India 20-23 October,1997.

Snedecor G. W. and Cochran W. G. 1989. Statistical Methods. Eighth Edition.Iowa State University Press, Ames, Iwoa 50010.

Zeileis A., Grothendieck G., Ryan J. A., Ulrich J. M., Andrews F.2020. Package ‘Zoo’.version 1.8-8. (URL: http://zoo.R-Forge.R-project.org).