



**HAL**  
open science

## Profiles of dairy cows with different productive lifespan emerge together from multiple traits assessed at first lactation: the case of a grassland-based dairy system

Luciano Barreto Mendes, M. Coppa, J. Rouel, B. Martin, Bertrand Dumont, Anne Ferlay, C. Espinasse, Fabienne Blanc

### ► To cite this version:

Luciano Barreto Mendes, M. Coppa, J. Rouel, B. Martin, Bertrand Dumont, et al.. Profiles of dairy cows with different productive lifespan emerge together from multiple traits assessed at first lactation: the case of a grassland-based dairy system. *Livestock Science*, 2021, pp.104443. 10.1016/j.livsci.2021.104443 . hal-03146933

**HAL Id: hal-03146933**

**<https://hal.science/hal-03146933v1>**

Submitted on 10 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Profiles of dairy cows with different productive lifespan emerge together from multiple**  
2 **traits assessed at first lactation: the case of a grassland-based dairy system**

3

4 **L. B. Mendes<sup>\*1</sup>, M. Coppa<sup>§</sup>, J. Rouel<sup>\*</sup>, B. Martin<sup>\*</sup>, B. Dumont<sup>\*</sup>, A. Ferlay<sup>\*</sup>, C.**  
5 **Espinasse<sup>†</sup> and F. Blanc<sup>\*</sup>**

6

7 \* Université Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, F-63122 Saint-  
8 Genès-Champanelle, France.

9 <sup>§</sup>Independent researcher at the Université Clermont Auvergne, INRAE, VetAgro Sup, UMR  
10 Herbivores, F-63122 Saint-Genès-Champanelle, France.

11 <sup>†</sup>INRAE, Herbipôle, 63122, Saint-Genès-Champanelle, France.

12 <sup>1</sup>Corresponding author: Luciano Barreto Mendes; Université Clermont Auvergne, INRAE,  
13 VetAgro Sup, UMR 1213 Herbivores, F-63122 Saint-Genès-Champanelle, France; +33 (0)4  
14 73 62 46 22; luciano.mendes@inrae.fr.

15 **HIGHLIGHTS**

- 16 • A trade-off was observed among productive lifespan of cows and milk yield, fertility  
17 and body weight at first lactation;
- 18 • Contrasting productive lifespan profiles defined as long, average and short were  
19 characterized by differing reproductive performance and milk production traits **in** first  
20 lactation;
- 21 • Individual placement of cows into groups of long, average or short productive  
22 longevity can be correctly predicted;
- 23 • Studying traits of cows at first lactation enables early identification of cows that have  
24 a greater probability of living longer.

25

26 **ABSTRACT**

27 This study focuses on extensive grassland-based dairy systems, where cows may be exposed  
28 to various hazards (climatic, health...) and limiting environments (feeding conditions) over  
29 their life. In such livestock systems, the ability of cows to remain in the herd (i.e. lifespan)  
30 reveals their ability to cope with such perturbations and to continue meeting the farmer's  
31 expectations. Productive lifespan was thus considered as a proxy of robustness in dairy cows.  
32 We hypothesized that productive lifespan of dairy cows can be predicted by a set of life  
33 function traits recorded during first lactation. A dataset composed of variables describing the  
34 entire productive life of 185 dairy cows managed in an experimental farm was used to test our  
35 hypothesis. This dataset contained information on 32 variables related to productive and  
36 reproductive performances, health status, average daily dry matter intake, body weight and  
37 body condition score describing the first lactation of all multiparous dairy cows, together with  
38 their productive lifespan (difference between age at culling and age at first calving).  
39 Clustering and multi-trait profile analyses of dairy cows revealed the existence of trade-offs  
40 between the productive lifespan of dairy cows and milk productivity, fertility and body weight  
41 at first lactation. For instance, cows with good **functional** performance **(in terms of health and**  
42 **reproduction traits)** but moderate milk production during first lactation had longer productive  
43 lifespan, whereas cows with moderate milk production but poor reproductive performance  
44 had a shorter one. We applied linear discriminant (LD) **modelling technique** to predict to  
45 which productive lifespan (short, average, long) profile an individual cow would belong,  
46 based on life function **and production** traits measured during the first lactation. K-fold cross-  
47 validation (k=4) of the LD model yielded an accuracy of 89% and precision ranged between  
48 89 and 97%. Prediction quality of the cross-validated LD model suggests that our approach of  
49 classifying individual dairy cows is relevant and advantageous compared to classifications  
50 done per average cluster or profile. Studying traits of cows at first lactation enables early

51 identification of cows that have a greater probability of living longer and thus have a better  
52 robustness.

53 **Key words:** clustering of dairy cows, robustness, trade-offs between functional and  
54 productive traits, discriminant analysis, survival, longevity.

## INTRODUCTION

55

56 Recent studies have emphasized that robustness is a relevant characteristic of dairy cows that  
57 should be addressed in the future (Roche et al., 2018). This is particularly true in extensive  
58 grassland-based production systems, where cows are exposed to limiting conditions and  
59 fluctuating environments (Delaby et al., 2018) and are able to express their genetic potential  
60 only partially. When facing nutritional constraints or challenges, modern dairy cows are  
61 expected to show a trade-off between life functions (growth, survival, health, reproduction,  
62 and lactation) (Blanc et al., 2006; Friggens and Newbold, 2007). Such trade-offs result from  
63 breeding programs mainly focused on lactation and, as a consequence, milk yield per dairy  
64 cow has doubled in the last 40 years (Oltenucu and Broom, 2010). Such a priority given to  
65 milk production traits has been shown to be deleterious to other traits, such as fertility  
66 (Albarrán-Portillo and Pollott, 2013; Berglund, 2008; Pryce et al., 2004), overall health (Berry  
67 et al., 2011), welfare (Oltenucu and Broom, 2010) and longevity (Essl, 1998), and it has  
68 yielded genotypes that show better dairy performance and efficiency in optimal environments  
69 (a rich environment where individuals are allowed to fully express their genetic potential  
70 (VandeHaar et al., 2016)), but are more susceptible to challenging environments (where  
71 animals are exposed to scarcities or restrictions (Blanc et al., 2006)).

72 Some authors have suggested that profiles of cows defined from expression of trade-offs  
73 between life functions that play **an important role** in challenging situations are linked to robust  
74 properties (Ollion et al., 2016). Recently, Friggens et al. (2017) defined animal robustness as  
75 “the ability, in the face of environmental constraints, to carry on doing the various things that  
76 the animal has to do to favor its future ability to reproduce”. From this definition it clearly  
77 appears that robustness is linked to the ability of the animal to remain in the herd (not being  
78 culled or sold and staying alive) and to match the farmer’s expectations (Ollion et al., 2018).  
79 Thus, an integrative way to estimate animal robustness is by looking at longevity (Friggens et

80 al., 2017). Our study aimed to test the link between these two previous approaches to assess  
81 animal robustness, i.e. to connect longevity to the ways cows express trade-offs between life  
82 functions.

83 As is well known, the longevity of a dairy cow is ultimately determined by the culling rules  
84 applied by the farmer. Culling can be driven by economic reasons such as herd renewal needs,  
85 feed costs, market demand for milk production, or slaughter (Edwards-Callaway et al., 2019),  
86 but most often it is first determined by biological reasons when cows are not able to meet the  
87 farmer's expectations anymore or fail to match the requirements of the production system,  
88 such as production amount and quality, for instance. Biological reasons for culling may be  
89 quite diverse like **poor** udder health or conformation, low fertility, low production, poor  
90 **milking**, leg problems, diseases (including metabolic diseases), poor temperament, injuries or  
91 old age. For example, poor udder health and poor fertility were reported to be the main reason  
92 for culling in organic and conventional Swedish dairy herds, respectively (Ahlman et al.,  
93 2011).

94 Because culling decisions involve risks related to discarding a cow that would otherwise be  
95 productive for a few more lactations, or retaining in the herd an unproductive individual for  
96 another cycle, we considered that the earlier we could detect cows likely to have high versus  
97 short longevity the more helpful it would be for the farmer. As Adriaens et al. (2020) have  
98 recently shown that production and activity traits measured at first lactation have the potential  
99 to predict the cows' lifetime resilience ranking within a herd, we hypothesized that, within a  
100 herd, groups of cows that would express different trade-offs between life functions during  
101 first lactation will have different productive lifetime performances (i.e. different  
102 robustness). The first objective of the present study was to establish whether cows differing in  
103 productive lifespan have different ways to prioritize life functions at first lactation. This is an  
104 *a posteriori* hypothesis testing approach, by assessing the variability of all traits and allowing

105 intrinsic trade-offs to emerge. We then tested *a priori* if it would be possible to predict  
106 whether or not individual cows would belong to groups of dairy cows that are statistically  
107 similar, according to their expression of health, growth, body reserves, milk production and  
108 reproductive traits at first lactation.

## 109 **MATERIALS AND METHODS**

### 110 *Description of farming facilities and management*

111 The data used in this study were collected between 1996 and 2016 at the INRAE experimental  
112 farm of Marcenat (INRAE, UE 1414 Herbipôle, France; 45°18'21"N, 2°50'13"E; average  
113 altitude 1075 m, DOI: <https://doi.org/10.15454/1.5572318050509348E12>), in the Massif  
114 Central mountain area of central France. Yearly average air temperature is 11 °C, and the  
115 annual average precipitation is 725 mm, with a maximum during autumn and moderate  
116 summer droughts. Dairy cows were fed hay from natural pastures, haylage and concentrate  
117 during the indoors season (from November to mid-April, between 1.0 and 1.5 t<sub>DMI</sub>·animal<sup>-1</sup>  
118 ·year<sup>-1</sup> of concentrate) and grazed for the rest of the year.

119 Housing facilities consist of a naturally ventilated barn subdivided into cubicles, with a  
120 maximum capacity of 110 cows. Storage of raw data [milk production and composition, body  
121 weight (BW), body condition score (BCS), fertility and health events] was performed using  
122 the data collection and management software Datalogic Aladdin™ (Datalogic S.p.A,  
123 Bologna, Italy). Milking was done year-round inside the barn with an automatic milking  
124 parlor twice a day (at 06:00 and 16:00). During the indoor season, forage was offered  
125 individually twice a day, and records of forage intake were obtained on an individual basis  
126 (forage offered minus refusal weighed three/four days a week). The amount of concentrate  
127 that was offered to the cows on a given week was determined to cover their nutritional  
128 requirements as quantified by the INRA feeding system (INRA, 1988) based on their milk



129 production **in** the previous week. Concentrate intake was monitored on an individual basis  
130 using an automatic feeder (Feeding Station FSC-40 DeLaval, Elancourt, France).  
131 Calving strategy was mostly seasonal, from cows artificially inseminated during the winter  
132 months. After a voluntary waiting period of at least two months, natural occurrence of heat (or  
133 *estrus*) was detected visually by trained staff in charge of the cows. Cases of *anestrus* were  
134 only dealt with for heifers, which were treated with hormones. Then, an artificial insemination  
135 campaign was conducted on the farm from November to February (4 months). Conception  
136 date was estimated from the size of the fetus obtained via ultrasound tests, and was further  
137 confirmed by comparing this date with the closest date of an observed heat event or an  
138 artificial insemination event. Cows that did not show signs of heat or were not confirmed as  
139 pregnant at the beginning of spring had a chance to be naturally impregnated by a bull, which  
140 was present in the herd during the entire grazing period. If a given cow was not confirmed  
141 pregnant at the end of the grazing/reproductive period, that cow was culled.

#### 142 ***Animal data***

143 Data on dairy performance, BW and BCS, fertility and health events along with feeding were  
144 collected at first lactation, and culling dates and reasons were recorded at the end of the  
145 productive career. Individual milk yield (MY, kg·d<sup>-1</sup>) was recorded daily, with the use of  
146 recording jars (DeLaval®, Ghent, Belgium) which rested on loading cells. Milk protein (MP,  
147 g·kg<sup>-1</sup>) and fat (MF, g·kg<sup>-1</sup>) contents, were analyzed weekly during 4 consecutive milkings  
148 (two consecutive days) using an infrared method (Milkoscan 4000, Foss System, Hillerød,  
149 Denmark). All cows were weighed with an electronic scale (Mettler-Toledo SAS, model  
150 Mettler Spider 2, Viroflay, France) after the morning milking for two consecutive days at  
151 least once a month. BCS was estimated through palpation using a scale from 0 to 5 points,  
152 with 0.25 interim steps (Bazin et al., 1984), by two trained staff members at least once a  
153 month. For each cow, reproductive status at first lactation was assessed by accounting for the

154 total number of heat events ( $\text{heat}_{\text{count}}$ ), the interval between calving and first heat ( $\text{int}_{\text{calving-heat}}$ ,  
155 d), the total number of artificial inseminations or natural inseminations by a bull ((AI +  
156 NI) $_{\text{count}}$ ), the number of successful inseminations ( $\text{conception}_{\text{count}}$ ) and the interval from  
157 calving to conception also commonly known as ‘days open’ (DO, d). In this study, the  
158 variable  $\text{conception}_{\text{count}}$  also included possible abortions that the cow had during the lactation.  
159 The number of natural inseminations was defined by adding up the total number of times  
160 when a heat event was observed at the field by a staff member (under the assumption that the  
161 cow **mated with** the bull every time a heat event happened) from turnout (start of natural  
162 breeding) to confirmation of pregnancy.  
163 Birthdate, age at first calving ( $\text{age}_{\text{1st calv}}$ , d) age at culling ( $\text{age}_{\text{culled}}$ , d), and reason for culling  
164 were tracked (more information on culling rules is provided below). Productive lifespan (d)  
165 was derived by the difference between total lifespan and age at first calving. Health status of  
166 cows was assessed throughout first lactation by recording incidences of mastitis ( $\text{mastitis}_{\text{count}}$ )  
167 and lameness ( $\text{lameness}_{\text{count}}$ ), as well as by counting any additional medical incidences and/or  
168 interventions ( $\text{med.int}_{\text{count}}$ ), such as emergency aid for accidents, other infections and diseases.  
169 Indicators of metabolic diseases were assessed by the percentage of time when milk fat  
170 content to milk protein content ratio ( $\%T_{\text{MF/MP}}$ ) was  $\leq 1.0$  and  $\geq 1.4$ , as expressing potential  
171 subclinical acidosis and ketosis, respectively (Nozière et al., 2018). Finally, feeding at the  
172 individual level was estimated during the winter as the sum of daily intakes of forage and  
173 concentrate ( $\text{kg DM}\cdot\text{d}^{-1}$ ), and the mean per week of lactation was calculated. It was not  
174 possible to record this variable when cows were grazing. All variables used in this study are  
175 presented in Table 1 along with their definitions, formulas and units.

### 176 ***Herd composition and culling strategy***

177 Data used in this study included individuals for which both first calving and end of the last  
178 lactation occurred during the study period. Only cows culled at the second lactation or later

179 were kept. Outliers were defined as cows with productive performance levels and calving-to-  
180 calving intervals out of the interquartile range  $\pm 1.5$  and then were removed. This yielded a  
181 total of 185 cows (114 Montbéliarde (MO) and 71 Prim'Holstein (HO)) that were used for  
182 further analysis.

183 Reasons for culling were: (1) low fertility (including empty, abortion, late calving and calving  
184 troubles); (2) poor health (including mastitis, lameness, acidosis, other infections and overall  
185 health status); (3) poor production performance, (4) herd renewal (including old age), and (5)  
186 accidents on farm.

### 187 *Statistical analyses*

188 All analyses in this study were performed with the statistical software R<sup>®</sup> (version 3.5.2, R  
189 Development Core Team, 2020).

190 With the aim of reducing the number of variables to perform a principal component analysis  
191 (PCA), a first variable selection was done to avoid variance inflation due to collinearity, i.e., a  
192 multicollinearity test was performed among selected variables. We used the variance inflation  
193 factor (VIF) as criterion to judge the presence of multicollinearity. The variable with the  
194 highest VIF was excluded, and new VIFs were calculated again. These steps were repeated  
195 until only variables that had a  $VIF \leq 10.0$  remained (Hair et al., 2014). All remaining  
196 variables were then normalized and scaled (between -1 and 1) prior to the analysis steps  
197 described below in order to avoid the variance inflated effect due to heterogeneous scaling of  
198 variables. Using the R<sup>®</sup> package FactoMineR v2.3 (Husson et al., 2020), we then performed  
199 PCA to determine which variables accounted for most of the variability found among  
200 individuals. With this, variables were further omitted by using a stepwise suppression  
201 criterion of significant correlation between a variable with any of the first three principal  
202 components. Only variables showing a significant  $r \geq 0.50$  ( $P \leq 0.05$ ) with at least one given  
203 component were kept.

204 Once identified through PCA, cluster analysis of individuals was performed using the  
205 Agglomerative Hierarchical Clustering (AHC) technique. The decision regarding the  
206 optimum number of clusters was based on an analysis of within-cluster variance, by detecting  
207 the minimum number of clusters after which an abrupt change in variance occurred.

208 Then, each variable (at original scale) was submitted to comparison tests in two steps: (1) the  
209 difference between the mean of each cluster and the respective mean obtained from the entire  
210 experimental herd was tested. For this process, the one-sample two-sided t-test was used for  
211 normally distributed variables, and the one-sample Wilcoxon signed rank test was applied to  
212 other variables; (2) between-cluster comparisons were performed using one-way ANOVA  
213 (considering cluster as a factor) for normally distributed variables and the Kruskal-Wallis test  
214 for the rest, followed by their respective post-hoc tests (Tukey and Wilcoxon signed rank  
215 tests, respectively).

216 We performed a trade-off analysis of groups of cows in each cluster by biologically  
217 interpreting the significant trait deviations (according to the cluster mean compared with the  
218 experimental herd mean) in each cluster. A trade-off was identified in a cluster when at least  
219 two traits presented statistically significant deviations with an opposite sign. When a trade-off  
220 occurred, we quantitatively profiled the group of individuals of that cluster according to the  
221 nature of the observed trade-off. Finally, we interpreted deviations in productive lifespan as a  
222 consequence of the combination of other traits and their trade-offs. The multi-trait profiling  
223 method used in this study was inspired by the approach proposed by Ollion et al. (2016). For  
224 illustrative purposes, one paragon (representative cow) per cluster was selected based on two  
225 criteria: (1) the shortest geometrical distance (calculated with individual coordinates obtained  
226 from PCA) of the individual to the cluster barycenter, and (2) the individual  
227 representativeness of cluster characteristics, i.e. whether or not paragons resemble average  
228 individuals per cluster.

229 The dependency of contingents of cows per cluster were compared to the contingents per  
230 reason for culling (poor fertility, poor health and other causes including low milk yield, old  
231 age, herd renewal and accidents), using Fisher's exact test.

232 The ability to discriminate cows according to their lifespan profile obtained in the previous  
233 step was tested through discriminant analysis. We used the package 'psych' v2.0.9 R<sup>®</sup>  
234 (Revelle, 2020) to run our linear discriminant (LD) regression, using the cluster classification  
235 obtained from PCA as our dependent variable.

236 We first calibrated the LD model with the entire available dataset, and then applied the k-fold  
237 cross-validation method (k=4), also with the aid of the R<sup>®</sup> package 'psych' v2.0.9. The quality  
238 of both calibrated and cross-validated models was further assessed through sensitivity  
239 (proportion of correctly identified individuals per class, or true positives), specificity  
240 (proportion of individuals that are correctly identified as not belonging to a specific class, or  
241 true negatives), precision (sensitivity divided by the number of true negatives) and accuracy  
242 (overall ability of the model to predict class correctly), as defined by Fawcett (2006).

243 In this study, results from all statistical analyses were considered significant when  $P \leq 0.05$ .

## 244 **RESULTS**

245 Table 2 provides results for variables describing milk production, BW, BCS, health status,  
246 fertility and feeding at first lactation, and longevity traits. Although we initially kept the  
247 "breed" variable in our statistical analyses, it did not turn out to be significant. For that reason,  
248 data per MO and HO were pooled together, and breed effect was not further considered in this  
249 study.

250 Descriptive summary statistics on milk production, BW, BCS, health status, fertility and  
251 feeding at first lactation are provided in Table 3.

252 The main reason that we generated as many as 32 variables (Table 1) to compose our dataset  
253 was to increase the chances that, after a variable selection process, at least one variable

254 representative of the most important life functions would remain. Indeed, 15 variables  
255 eventually remained after PCA (Table 3), representing **the main** life functions, as defined by  
256 Friggens and Newbold (2007): survival (productive lifespan), **growth** ( $BW_{calving}$ ,  $BW_{loss}$ ),  
257 **body reserves** ( $BCS_{min}$ ), reproduction ( $heat_{count}$ ,  $(AI+NI)_{count}$ ,  $int_{calving-1st\ heat}$ ,  $DO$ ), **and**  
258 **lactation** (through productive traits  $\overline{MY}_{305d}$ ,  $cumulativeMY_{peak}$ ,  $cumulativeMY_{4wk}$ ,  
259 **Persistency**,  $\overline{MP}$ ), plus feeding ( $\overline{concentrate}_{10wk}$ ).

260 Health variables were eliminated in the selection process for PCA to avoid inflation problems  
261 and do not appear in Table 3. However, they were considered again when performing profile  
262 analysis and LD regression, where inflation is not an issue.

### 263 *PCA and clustering*

264 Through the variable reduction process involved in PCA, we obtained 15 variables from the  
265 initial 32 variables considered in this study (Table 1). A total of four principal components  
266 (PC) with eigenvalues  $\geq 1$  accounted for 68% of the total variability of the dataset (Table 3),  
267 concerning milk yield, BW, feeding and reproduction.

268 Four variables concerning milk yield ( $\overline{MY}_{305d}$ ,  $cumulativeMY_{peak}$ ,  $cumulativeMY_{4wk}$ ,  
269 persistency) together with  $int_{dry}$  had the highest correlations with PC1 as well as  
270 ( $\overline{concentrate}_{10wk}$ ). All these variables, except persistency were positively correlated with  
271 PC1. Productive lifespan variable was most strongly (and negatively) correlated with PC2 (-  
272 0.34,  $P < 0.05$ ), while  $heat_{count}$ ,  $(AI+NI)_{count}$  and  $DO$  were positively correlated with this  
273 component (0.69, 0.75 and 0.70, respectively). PC3 had the highest contributions from  $\overline{MP}$ ,  
274  $BW_{calving}$ ,  $BW_{min}$  and  $BCS_{min}$  (positively correlated with PC3, with correlations 0.54, 0.73,  
275 0.77 and 0.59, respectively, all with  $P < 0.0001$ ). Finally, PC4 had a strong positive correlation  
276 with  $int_{calving-1st\ heat}$  (0.68,  $P < 0.0001$ ).

277 PC1 distinguished individuals that presented high levels of milk yield at first lactation  
278 ( $\overline{MY}_{305d}$ ) associated with high consumption of concentrate ( $\overline{concentrate}_{10wk}$ ), long times to

279 successful reproduction ( $\text{int}_{\text{calving}-1^{\text{st}} \text{heat}}$  and  $\text{int}_{\text{calving}-\text{conception}}$ ) and also a long dry  
280 period between first and second lactation. PC1 also distinguished individuals who have a high  
281 peak ( $\text{cumulativeMY}_{\text{peak}}$ ), a considerable increase in milk yield up to the first four weeks of  
282 lactation ( $\text{cumulativeMY}_{4\text{wk}}$ ) and a weak persistency.

283 PC2 distinguished cows with low milk yield at first lactation ( $\overline{\text{MY}}_{305\text{d}}$ ) associated with more  
284 counts of heat ( $\text{heat}_{\text{count}}$ ) and insemination attempts ( $(\text{AI} + \text{NI})_{\text{count}}$ ). According to PC2,  
285 cows characterized by lower milk yield and difficulties in conception also presented lower  
286 survival (productive lifespan).

287 PC3 separated cows according to high protein content in milk ( $\overline{\text{MP}}$ ) and their high ability to  
288 maintain body reserves ( $\text{BW}_{\text{calving}}$ ,  $\text{BW}_{\text{min}}$  and  $\text{BCS}_{\text{min}}$ ). Finally, PC4 mainly distinguished  
289 cows that presented longer times to restart estrus ( $\text{int}_{\text{calving}-1^{\text{st}} \text{heat}}$ ).

290 From the AHC results, individual cows were differentiated into three clusters. The  
291 partitioning of cows per cluster was 80, 28 and 77 individuals for clusters 1, 2 and 3,  
292 respectively.

293 Cluster 1 was characterized by cows that had a productive lifespan that was 11% longer than  
294 the experimental herd mean. They also produced 10 to 24% (for  $\overline{\text{MY}}_{305\text{d}}$  and  
295  $\text{cumulativeMY}_{\text{peak}}$ , respectively) less milk at first lactation than the average. The average  
296 lengths of lactation and the dry period were significantly 5 and 12% shorter for cows in  
297 cluster 1, as compared to the herd mean, respectively. Cows from cluster 1 had a  $\text{BW}_{\text{calving}}$  2%  
298 lower than the herd mean, but a  $\text{BCS}_{\text{min}}$  8% higher. Cluster 1 was characterized by cows that  
299 had  $\%T_{\text{MF/MP}>1.4}$  and  $\text{lameness}_{\text{count}}$  30 and 44% lower than the herd mean, respectively. These  
300 cows were more fertile than the herd mean ( $\text{heat}_{\text{count}}$  and  $(\text{AI}+\text{NI})_{\text{count}}$  decreased by 21 and  
301 28%, respectively) and became pregnant earlier than the herd mean, as indicated by shorter  
302  $\text{int}_{\text{calving}-1^{\text{st}} \text{heat}}$ ,  $\text{int}_{\text{calving}-1^{\text{st}} \text{insemination}}$ ,  $\text{int}_{\text{calving}-\text{conception}}$ ,  $\text{int}_{1^{\text{st}} \text{heat}-\text{conception}}$  and CI

303 (6 to 23% shorter than the herd mean). Concentrate consumption was significantly lower in  
304 this cluster compared to the herd mean (-16%).

305 Cluster 2 was characterized by cows that had a productive lifespan 27% shorter than the herd  
306 mean. Average milk production of these primiparous cows was, in general, similar to that of  
307 the herd mean, except for variable  $\overline{MY}_{305d}$ , which was reduced by 13% compared to the herd  
308 mean. Moreover, cows from cluster 2 had significantly longer DIM (+16%) which was  
309 followed by a longer dry period ( $int_{dry}$ , +29%) than the herd mean. They took a longer time to  
310 express first heat after calving ( $int_{calving - 1st\ heat}$ , +27%), had lower fertility than the herd mean,  
311 presenting more  $heat_{count}$  and  $(AI + NI)_{count}$  (76% and 91%, respectively) and finally showed  
312 longer DO than the herd mean (+59%). Health variables observed in this cluster did not differ  
313 from the herd mean.

314 Cows from cluster 3, which was characterized by a productive lifespan that did not differ  
315 from the herd mean, presented better milk production performance in relation to the herd  
316 mean, by presenting higher  $\overline{MY}$ ,  $\overline{MY}_{305d}$ ,  $cumulativeMY_{305d}$ ,  $cumulativeMY_{peak}$ ,  
317  $cumulativeMY_{4wk}$  and  $cumulativeMY_{24wk}$  (increased by between 14 and 27%). They showed  
318 lower persistency of milk production than the herd mean. These cows also presented  $BW_{calving}$   
319 and  $BW_{min}$  that were significantly higher than the herd mean by 3 and 2%, respectively.  
320 Concerning health variables, cows in cluster 3 presented 11% lower  $\%T_{MF/MP < 1.0}$ , but  
321 presented increased  $mastitis_{count}$  and  $lameness_{count}$  (30 and 35%, respectively) compared to the  
322 herd mean. They presented an  $int_{calving - 1st\ heat}$  14% higher than the herd mean, but an  $int_{1st\ heat -$   
323  $conception}$  18% shorter, resulting in an  $int_{calving - conception}$  similar to the herd mean. Cows in  
324 this cluster ate more concentrate than the herd mean (+18%).

### 325 ***Reasons for culling per cluster***

326 The majority of cows in the herd were culled due to poor fertility (59%) or poor health (26%)  
327 and only 15% were culled due to other reasons like poor milk production, herd renewal, old



328 age or accidents. This is further confirmed by the contingents of cows per cluster and reasons  
329 for culling that are presented in Table 5. Poor fertility was the leading reason for culling cows  
330 in clusters 1, 2 and 3, but was more frequent in cluster 2 (49%, 75%, 64% respectively). The  
331 **second reason for culling was** poor health with 33%, 25% and 22%, respectively. Culling for  
332 reasons other than reproduction or health was observed in clusters 1 and 3 (18% and 15%,  
333 respectively), but not in cluster 2.

### 334 *Discrimination of cows according to clustered profiles*

335 Results from LD analysis revealed that between-class variance explained by linear  
336 discriminants 1 and 2 was 56 and 44%, respectively. Figure 1 illustrates how individual cows  
337 per cluster are distributed in a Cartesian plane which has LD1 and LD2 as orthogonal axes.  
338 Figure 2 depicts the variables' contributions to LD1 and LD2. The main positive contributions  
339 to LD1 came from  $\overline{MY}_{305d}$ , DO,  $\overline{concentrate}_{10wk}$  and  $(AI+NI)_{count}$ , while the main negative  
340 contributions to this LD came from persistency and cumulativeMY<sub>4wk</sub>.

341 The main positive variable contribution to LD2 came from  $\overline{MY}_{305d}$ , while the main negative  
342 contributions to this LD were from variables  $(AI+NI)_{count}$ ,  $int_{dry}$ ,  $heat_{count}$ , persistency and DO.  
343 Figure 1 suggests that LD1 discriminates clusters 1 and 3 better, while LD2 discriminates  
344 cluster 2 from both clusters 1 and 3 better.

345 Between-cluster comparison tests (Table 4) confirm the effect of LD1 and LD2 in  
346 discriminating clusters. In other words, means for clusters 1 and 3 were significantly different  
347 for the main variables that describe LD1 ( $\overline{MY}_{305d}$ ,  $\overline{concentrate}_{10wk}$ , Persistency,  
348 CumulativeMY<sub>4wk</sub> and  $(AI+NI)_{count}$ ). Means from cluster 2 were significantly different from  
349 those of clusters 1 and 3 for the variables that mostly contribute to LD2 (mainly  $heat_{count}$ ,  
350  $(AI+NI)_{count}$ ,  $int_{dry}$  and DO).

351 Results from cross-validation of the LD model indicate an overall accuracy of 86%. Other  
352 model quality measures are presented in Table 6.

353 **DISCUSSION**

354 *Characterization of trade-offs and profiling clusters of dairy cows*

355 Cluster analysis revealed different strategies of individual cows regarding priority given to  
356 lactation, reproduction and health performance at first lactation that ultimately were  
357 associated with significant differences in productive lifespan. A trade-off between milk  
358 production and functional traits (reproduction and health) at first lactation existed in cluster 1  
359 in parallel to a trend to higher productive lifespan. Thus, productive lifespan was observed to  
360 be longer for cows in cluster 1 in which reproduction, health and BCS safeguarding were  
361 prioritized to the detriment of milk production at first lactation. If we suppose that such  
362 priorities remain in successive lactations, the higher longevity observed in this cluster is not  
363 surprising as reproduction and health have been reported to be the main reasons for culling in  
364 dairy herds (Ahlman et al., 2011; De Vries and Marcondes, 2020). Other studies have  
365 reported positive correlations between longevity and reproduction traits (De Vries and  
366 Marcondes, 2020; O’Sullivan et al., 2020), good health (Alvåsen et al., 2018; Shabalina et al.,  
367 2020) and ability to preserve body reserves (Zavadilová et al., 2011). Jimenez-Krassel et al.  
368 (2015) found that young adult dairy heifers with higher concentrations of anti-Müllerian  
369 hormone (which is positively associated with follicle count, ovarian function, and fertility)  
370 had a longer productive lifespan. O’Sullivan et al. (2020) reported that survival of Holstein-  
371 Friesian cows in pasture-based systems has been significantly improved in the genetic  
372 divergent group of cows presenting the highest Irish Economic Breeding Index (EBI)  
373 compared to the average. These “top” EBI cows achieved greater pregnancy rates, presented  
374 fewer services per cow and shorter intervals between mating start date and pregnancy  
375 compared to the average, and, like cows in cluster 1, they maintained **higher** mean BCS  
376 during lactation. Our results are also in agreement with those of Shabalina et al. (2020), who  
377 estimated negative genetic correlations between longevity and occurrence of common health

378 disorders in German herds of Holstein cows. Furthermore, Alvåsen et al. (2018) found that  
379 Swedish farmers who received more visits from the veterinarian had dairy herds with shorter  
380 longevity.

381 Cows in cluster 2 succeeded in conceiving at first lactation (as all cows considered in this  
382 study), but with a large delay compared to the herd mean (+48 days). Consequently, these  
383 cows re-calved with a delay of more than 2 months (+72 days for CI) compared to the herd  
384 mean and started their subsequent breeding season later. In a pasture-based system, the  
385 reproductive season is often time-bound in order to make lactation match grass availability  
386 (Delaby et al., 2018). In other words, seasonal or grouped calving derive from the strong  
387 constraints that characterize pasture-based dairy farms, with a reproductive season of only 3  
388 months in order to synchronize the beginning of lactations with the period of high availability  
389 of forage (Roche et al., 2018). Consequently, cows of cluster 2 had less chance of becoming  
390 pregnant at the second and subsequent lactations within the period defined by the farmer, first  
391 because their breeding season is shortened compared to the herd mean and second because  
392 they need more time to conceive. This is in agreement with the reasons for culling observed in  
393 this cluster with 75% of cows culled because of reproduction issues. Reimus et al. (2018) also  
394 showed that Estonian dairy cows with lower conception rate presented higher mortality, i.e.  
395 unassisted death and euthanasia. No trade-off between reproduction and any other life  
396 function was observed in our study for cluster 2, except that they showed slightly higher milk  
397 solid contents (MF and MP) than the herd mean. In fact, Bedere et al. (2017a) found that,  
398 when underfed, high-yielding cows that presented higher milk solid content (MF + MP) were  
399 more prone to fertility failure, which might explain why they presented conception delays.

400 When reproduction is not a prioritized life function, the risk for the cow is lower longevity  
401 and for the farmer it is inability to use voluntary culling basing on economic reasons. No cows  
402 in cluster 2 were culled for reasons other than biological ones (poor fertility or poor health).

403 Cows in cluster 3 were able to manage high levels of milk yield regarding no trade-off with  
404 reproduction as they succeeded in conceiving with no delay compared to the herd mean. Such  
405 an ability to maintain reproduction was observed despite a higher calving to first heat interval  
406 (+10 days) and has also been reported in high-producing cows by Bedere et al. (2017b), who  
407 considered this to be due to delays in commencement of luteal activity and first service.  
408 However, the productive lifespan of these cows did not differ from the herd mean, probably  
409 because of high milk yield traded-off with health. Higher occurrence of mastitis and lameness  
410 observed in this cluster at first lactation may have increased their culling risk due to poor  
411 health, as **also** described by Shabalina et al. (2020) for German Holstein cows. However, the  
412 proportion of cows culled for poor health reasons was not higher for cows in cluster 3 (22%)  
413 than for those in clusters 1 (33%) and 2 (25%), which probably means that greater health  
414 issues observed at first lactation for cows of cluster 3 would have not been amplified for  
415 subsequent lactations.

416 The difference in concentrate intake ( $\overline{\text{concentrate}}_{10 \text{ wk}}$ ) that was observed between clusters  
417 is explained by the fact that profiles were characterized by different average levels of milk  
418 production. In our study, the feeding system used to define amounts of concentrate to be  
419 offered per cow and per day was adjusted to the cow's milk yield the week before. The fact  
420 that we did not identify any difference in forage intake between clusters is probably because  
421 our analysis considered data from first lactation only. Grandl et al. (2016) observed that the  
422 average organic dry matter intake of dairy cows increased with age.

423 Profiling reveals that high productive lifespan is mainly associated with reproductive success.  
424 Poor fertility and poor health were the main reasons why cows were ultimately culled from  
425 the experimental herd. In this study, culling due to old age or economic reasons was **present**  
426 in clusters 1 and 3, but not in cluster 2. Cows culled for old age or herd renewal are probably  
427 the most robust cows because they went on “doing what they have to do” throughout their

428 lives, without showing reproductive or health problems exceeding the tolerance threshold  
429 defined by the farmer. The dependence we found between contingents of cows per reason for  
430 culling (reproduction, health and economic) and per productive longevity profile (low,  
431 medium and high) indicates that the individuals that presented unfavorable milk production  
432 and reproductive or health traits at first lactation were ultimately culled for that same reason.  
433 For instance, cows from the low productive longevity profile (cluster 2) were mostly culled  
434 due to reproductive issues, while they also presented the poorest fertility at first lactation.  
435 These outcomes sustain the approach tested in this study and already suggested by Adriaens et  
436 al. (2020), that productive longevity ranking within a herd can be predicted by phenotyping  
437 several traits at first lactation. What our study adds to this idea is that clustering cows at first  
438 lactation from the way they express life functions could also help to predict the reason for  
439 their culling.

440 **Study results are bound to grassland-based dairy farming systems with seasonal calving,**  
441 where reproduction represents is a major cause of culling. This is particularly the case in  
442 France, Ireland and New Zealand, which is non-negligible. In other parts of the world it seems  
443 that health is the major cause for culling, while reproduction comes in second place (De Vries  
444 and Marcondes, 2020).

445 ***Profiling of cows at first lactation with low, medium and high productive lifespan is***  
446 ***predictable***

447 Clustering and profiling of cows allowed us to observe the trade-offs between productive  
448 lifespan, productive and functional traits *a posteriori*, i.e., without previously hypothesizing  
449 that productive lifespan is explained by such trade-offs. Conversely, the LD model obtained in  
450 this study stems from an *a priori* testing of the hypothesis that productive longevity and  
451 functional traits measured at first lactation are linked.

452 Discrimination by LD1 and LD2 of groups of cows agrees well with the profile analysis  
453 performed above. Although the variable productive lifespan was not directly present in the  
454 discriminant analysis, it was indirectly represented in the classification of clusters. This was  
455 clearly seen by taking a look at the selected individual paragons (Figure 3), where all  
456 individuals had distinct productive lifespans.

457 The main contributions to LD1 and LD2 are shared between milk production and  
458 reproduction traits, which have opposing or polarizing effects. A distinction should be made  
459 between production variables that describe the entire lactation cycle, such as  $\overline{MY}_{305d}$ , and  
460 variables that describe the geometry or shape of the lactation curve, such as persistency and  
461 cumulative  $MY_{4wk}$ . Indeed, two cows with identical  $\overline{MY}_{305d}$  may have lactation curves of  
462 different shapes.

463 The polarizing effect seen in LD1 (high overall productivity, longer lactation and poorer  
464 fertility, as opposed to high productivity at the beginning of lactation, with a more  
465 accentuated drop in the lactation curve after the peak) is supported by Muir et al. (2004), who  
466 showed that primiparous cows which had more calving or conception difficulties also  
467 presented more persistent first lactation. The same idea is further supported in this study, as  
468 illustrated by paragons for clusters 1 and 3 (Figure 3).

469 The inverse production vs. reproduction relationship comprised in LD2 is well known and has  
470 been mentioned before in several studies on dairy cows, such as Nebel and McGilliard (1993),  
471 Leroy et al. (2008) and Pires et al. (2015). Further evidence in this study that explains the  
472 effect of LD2 is the contrast between the paragons for cluster 2 with longer lactation and  
473 higher milk yield associated with greater (AI+NI) counts and longer dry period, and the  
474 paragons from clusters 1 and 3 (Figures 3A and 3C, respectively).

475 Concerning the predictive quality of the LD model, its overall ability to distinguish  
476 individuals from clusters 1 and 2 (the clusters with high and low productive lifespan,

477 respectively) with 97 and 86% correct classification is considered good for calibrated and  
478 cross-validated models, respectively. This shows that despite the fact that the variable  
479 productive lifespan was not directly included in the model, its influence on clustering  
480 individuals was significant.

481 The discriminant analysis with cross-validation suggests that our approach is relevant for  
482 moving from the average level per cluster/profile to the individual level in order to classify  
483 the animals. Our good results in cross-validation reinforce the relevance of the grouping  
484 approach provided by LD analysis and the effectiveness of the selected variables in predicting  
485 productive lifespan.

## 486 CONCLUSIONS

487 In this study, we revealed a clear association between productive longevity and trade-offs  
488 between life functions (milk production, reproduction, BW or body reserves lability and  
489 health) assessed at first lactation. Such an association was observed irrespective of the **two**  
490 **breeds used in this study (MO and HO)** and for cows managed in a grassland-based dairy  
491 farm. Cows with good reproductive performance but moderate milk production during first  
492 lactation had a longer productive lifespan, whereas cows with moderate milk production but  
493 poor reproductive performance had a shorter one. Productive lifespan can be predicted from a  
494 trade-off strategy at an individual level through discriminant analysis, by correctly classifying  
495 the group membership of individuals. Key variables to predict lifespan group membership  
496 were the overall milk production at first lactation, **traits characterizing the shape of the**  
497 **lactation curve** and fertility traits.

498 Our approach allows the prediction of longevity ranking of cows from traits measured at first  
499 lactation and provides a means for the early identification of those cows that have a greater  
500 probability of living longer and thus for the selection of robust cows within the herd.

## 501 ACKNOWLEDGEMENTS

502 This research received partial funding from GenTORE Project, a European Union H2020  
503 Research and Innovation Program, under agreement No. 727213, and falls within the thematic  
504 area of the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25). The  
505 authors would like to acknowledge Philippe Pradel and Matthieu Bouchon for their aid  
506 retrieving some of the data used in this study. Final thanks go to the anonymous reviewers for  
507 their helpful comments on earlier versions of the manuscript.

## 508 REFERENCES

- 509 Adriaens, I., Friggens, N.C., Ouweltjes, W., Scott, H., Aernouts, B., Statham, J., 2020.  
510 Productive life span and resilience rank can be predicted from on-farm first-parity sensor  
511 time series but not using a common equation across farms. *J. Dairy Sci.*  
512 <https://doi.org/10.3168/jds.2019-17826>
- 513 Ahlman, T., Berglund, B., Rydhmer, L., Strandberg, E., 2011. Culling reasons in organic and  
514 conventional dairy herds and genotype by environment interaction for longevity. *J. Dairy*  
515 *Sci.* 94, 1568–1575. <https://doi.org/10.3168/jds.2010-3483>
- 516 Albarrán-Portillo, B., Pollott, G.E., 2013. The relationship between fertility and lactation  
517 characteristics in Holstein cows on United Kingdom commercial dairy farms. *J. Dairy*  
518 *Sci.* 96, 635–646. <https://doi.org/10.3168/jds.2012-5632>
- 519 Alvåsen, K., Dohoo, I., Roth, A., Emanuelson, U., 2018. Farm characteristics and  
520 management routines related to cow longevity: a survey among Swedish dairy farmers.  
521 *Acta Vet. Scand.* 60, 38. <https://doi.org/10.1186/s13028-018-0390-8>
- 522 Bazin, S., Augéard, P., Carteau, M., Champion, H., Chilliard, Y., Cuyllé, G., Dissenhaus, C.,  
523 Durand, G., Espinasse, R., Gascoin, A., Godineau, M., Jouanne, D., Olivier, O.,  
524 Remond, B., 1984. Grille de notation de l'état d'engraissement des vaches pie noires.  
525 *Inst. Tech. l'Élevage Bovin*, Paris, France.
- 526 Bedere, N., Disenhaus, C., Ducrocq, V., Leurent-Colette, S., Delaby, L., 2017a. Ability of



527 dairy cows to ensure pregnancy according to breed and genetic merit for production traits  
528 under contrasted pasture-based systems. *J. Dairy Sci.* 100, 2812–2827.  
529 <https://doi.org/10.3168/jds.2016-11588>

530 Bedere, N., Disenhaus, C., Ducrocq, V., Leurent-Colette, S., Delaby, L., 2017b. Ability of  
531 dairy cows to be inseminated according to breed and genetic merit for production traits  
532 under contrasting pasture-based feeding systems. *animal* 11, 826–835.  
533 <https://doi.org/10.1017/S1751731116002111>

534 Berglund, B., 2008. Genetic Improvement of Dairy Cow Reproductive Performance. *Reprod.*  
535 *Domest. Anim.* 43, 89–95. <https://doi.org/10.1111/j.1439-0531.2008.01147.x>

536 Berry, D.P., Bermingham, M.L., Good, M., More, S.J., 2011. Genetics of animal health and  
537 disease in cattle. *Ir. Vet. J.* 64, 5. <https://doi.org/10.1186/2046-0481-64-5>

538 Blanc, F., Bocquier, F., Agabriel, J., D’hour, P., Chilliard, Y., 2006. Adaptive abilities of the  
539 females and sustainability of ruminant livestock systems. A review. *Anim. Res.* 55, 489–  
540 510. <https://doi.org/10.1051/animres:2006040>

541 De Vries, A., Marcondes, M.I., 2020. Review: Overview of factors affecting productive  
542 lifespan of dairy cows. *animal* 14, s155–s164.  
543 <https://doi.org/10.1017/S1751731119003264>

544 Delaby, L., Buckley, F., McHugh, N., Blanc, F., 2018. Robust animals for grass-based  
545 production systems, in: *Proceedings of the 27th General Meeting of the European*  
546 *Grassland Federation*. TEAGASC, Cork, Ireland, pp. 389–400.

547 Edwards-Callaway, L.N., Walker, J., Tucker, C.B., 2019. Culling Decisions and Dairy Cattle  
548 Welfare During Transport to Slaughter in the United States. *Front. Vet. Sci.* 5.  
549 <https://doi.org/10.3389/fvets.2018.00343>

550 Essl, A., 1998. Longevity in dairy cattle breeding: a review. *Livest. Prod. Sci.* 57, 79–89.  
551 [https://doi.org/10.1016/S0301-6226\(98\)00160-2](https://doi.org/10.1016/S0301-6226(98)00160-2)

552 Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognit. Lett.* 27, 861–874.

553 Friggens, N.C., Blanc, F., Berry, D.P., Puillet, L., 2017. Review: Deciphering animal  
554 robustness. A synthesis to facilitate its use in livestock breeding and management.  
555 *animal* 1–15. <https://doi.org/10.1017/S175173111700088X>

556 Friggens, N.C., Newbold, J.R., 2007. Towards a biological basis for predicting nutrient  
557 partitioning: the dairy cow as an example. *animal* 1, 87.  
558 <https://doi.org/10.1017/S1751731107657772>

559 Grandl, F., Luzi, S.P., Furger, M., Zeitz, J.O., Leiber, F., Ortmann, S., Clauss, M., Kreuzer,  
560 M., Schwarm, A., 2016. Biological implications of longevity in dairy cows: 1. Changes  
561 in feed intake, feeding behavior, and digestion with age. *J. Dairy Sci.* 99, 3457–3471.  
562 <https://doi.org/10.3168/jds.2015-10261>

563 Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2014. *Multivariate Data Analysis*, 7th  
564 Editio. ed. Pearson, Edinburgh.

565 Husson, F., Josse, J., Le, S., Mazet, J., 2020. Package “FactoMineR”: Multivariate  
566 Exploratory Data Analysis and Data Mining.

567 INRA, 1988. *Alimentation des bovins, ovins & caprins*. Cornell University.

568 Jimenez-Krassel, F., Scheetz, D.M., Neuder, L.M., Ireland, J.L.H., Pursley, J.R., Smith, G.W.,  
569 Tempelman, R.J., Ferris, T., Roudebush, W.E., Mossa, F., Lonergan, P., Evans, A.C.O.,  
570 Ireland, J.J., 2015. Concentration of anti-Müllerian hormone in dairy heifers is positively  
571 associated with productive herd life. *J. Dairy Sci.* 98, 3036–3045.  
572 <https://doi.org/10.3168/jds.2014-8130>

573 Leroy, J., Van Soom, A., Opsomer, G., Goovaerts, I., Bols, P., 2008. Reduced Fertility in  
574 High-yielding Dairy Cows: Are the Oocyte and Embryo in Danger? Part II Mechanisms  
575 Linking Nutrition and Reduced Oocyte and Embryo Quality in High-yielding Dairy  
576 Cows\*. *Reprod. Domest. Anim.* 43, 623–632. <https://doi.org/10.1111/j.1439->

577 0531.2007.00961.x

578 Muir, B.L., Fatehi, J., Schaeffer, L.R., 2004. Genetic Relationships Between Persistency and  
579 Reproductive Performance in First-Lactation Canadian Holsteins. *J. Dairy Sci.* 87, 3029–  
580 3037. [https://doi.org/10.3168/jds.S0022-0302\(04\)73435-9](https://doi.org/10.3168/jds.S0022-0302(04)73435-9)

581 Nebel, R.L., McGilliard, M.L., 1993. Interactions of High Milk Yield and Reproductive  
582 Performance in Dairy Cows. *J. Dairy Sci.* 76, 3257–3268.  
583 [https://doi.org/10.3168/jds.S0022-0302\(93\)77662-6](https://doi.org/10.3168/jds.S0022-0302(93)77662-6)

584 Nozière, P., Sauvant, D., Delaby, L. (Eds.), 2018. Alimentation des ruminants: apports  
585 nutritionnels - besoins et réponses des animaux - rationnement - tables de valeurs des  
586 aliments. Editions Quae, Versailles, France.

587 O’Sullivan, M., Butler, S.T., Pierce, K.M., Crowe, M.A., O’Sullivan, K., Fitzgerald, R.,  
588 Buckley, F., 2020. Reproductive efficiency and survival of Holstein-Friesian cows of  
589 divergent Economic Breeding Index, evaluated under seasonal calving pasture-based  
590 management. *J. Dairy Sci.* 103, 1685–1700. <https://doi.org/10.3168/jds.2019-17374>

591 Ollion, E., Blanc, F., Ingrand, S., 2018. How livestock farmers define dairy cows. *Fourrages*  
592 235, 219–227.

593 Ollion, E., Ingrand, S., Delaby, L., Trommenschlager, J.M., Colette-Leurent, S., Blanc, F.,  
594 2016. Assessing the diversity of trade-offs between life functions in early lactation dairy  
595 cows. *Livest. Sci.* 183, 98–107. <https://doi.org/10.1016/j.livsci.2015.11.016>

596 Oltenacu, P.A., Broom, D.M., 2010. The impact of genetic selection for increased milk yield  
597 on the welfare of dairy cows. *Anim. Welf.* 19, 39–49.

598 Pires, J.A.A., Chilliard, Y., Delavaud, C., Rouel, J., Pomiès, D., Blanc, F., 2015.  
599 Physiological adaptations and ovarian cyclicity of Holstein and Montbéliarde cows under  
600 two low-input production systems. *animal* 9, 1986–1995.  
601 <https://doi.org/10.1017/S1751731115001317>

602 Pryce, J.E., Royal, M.D., Garnsworthy, P.C., Mao, I.L., 2004. Fertility in the high-producing  
603 dairy cow. *Livest. Prod. Sci.* 86, 125–135. [https://doi.org/10.1016/S0301-](https://doi.org/10.1016/S0301-6226(03)00145-3)  
604 [6226\(03\)00145-3](https://doi.org/10.1016/S0301-6226(03)00145-3)

605 R Development Core Team, 2020. R: A language and environment for statistical computing.  
606 R Foundation for Statistical Computing, Vienna, Austria. URL [https://www.R-](https://www.R-project.org/)  
607 [project.org/](https://www.R-project.org/).

608 Reimus, K., Orro, T., Emanuelson, U., Viltrop, A., Mõtus, K., 2018. On-farm mortality and  
609 related risk factors in Estonian dairy cows. *Prev. Vet. Med.* 155, 53–60.  
610 <https://doi.org/10.1016/j.prevetmed.2018.04.006>

611 Revelle, W., 2020. Package “psych”: Procedures for Psychological, Psychometric and  
612 Personality Research.

613 Roche, J.R., Berry, D.P., Delaby, L., Dillon, P.G., Horan, B., MacDonald, K.A., Neal, M.,  
614 2018. New considerations to refine breeding objectives of dairy cows for increasing  
615 robustness and sustainability of grass-based milk production systems. *animal* 12, s350–  
616 s362. <https://doi.org/10.1017/S1751731118002471>

617 Shabalina, T., Yin, T., König, S., 2020. Influence of common health disorders on the length of  
618 productive life and stayability in German Holstein cows. *J. Dairy Sci.* 103, 583–596.  
619 <https://doi.org/10.3168/jds.2019-16985>

620 VandeHaar, M.J., Armentano, L.E., Weigel, K., Spurlock, D.M., Tempelman, R.J.,  
621 Veerkamp, R., 2016. Harnessing the genetics of the modern dairy cow to continue  
622 improvements in feed efficiency. *J. Dairy Sci.* 99, 4941–4954.  
623 <https://doi.org/10.3168/jds.2015-10352>

624 Zavadilová, L., Němcová, E., Štípková, M., 2011. Effect of type traits on functional longevity  
625 of Czech Holstein cows estimated from a Cox proportional hazards model. *J. Dairy Sci.*  
626 94, 4090–4099. <https://doi.org/10.3168/jds.2010-3684>

627

628

## FIGURE CAPTIONS

629 (Mendes1)

630 Figure 1. Classification of individuals in three profiles according to a linear discriminant  
631 model.

632

633 (Mendes2)

634 Figure 2. Variable contribution to each linear discriminant (LD).

635

636 (Mendes3)

637 Figure 3. Milk yield (MY, kg.d<sup>-1</sup>) concerning the entire first lactation and beginning of the  
638 second lactation for three selected paragons of clusters 1, 2 and 3 (A, B and C, respectively).  
639 Vertical arrows represent natural and/or artificial insemination events. Horizontal arrows  
640 represent the length of dry period. Bar plots represent productive lifespan (d) per paragon.

641

642

644 Table 1. Definition of variables related to milk production, body weight, body condition  
 645 score, health status, fertility and feeding at first lactation and longevity traits, their equations  
 646 and units

Variable and equation	Definition and units
$\overline{MY} = \frac{1}{n} \sum_{i=1}^{i=n} MY_i$	Average daily milk yield (MY) for the entire lactation cycle (kg·d <sup>-1</sup> *); n is the number of weeks in a single lactation (1)
$\overline{MY}_{305d} = \frac{1}{44} \sum_{i=1}^{i=44} MY_{305d,i}$	Average daily milk yield (MY) for up to a standard lactation cycle of 305d (kg·d <sup>-1</sup> *); 44 is the considered number of weeks comprised in a standard lactation (2)
$DIM = 7 \cdot \sum_{i=1}^n i$	Time interval between calving and the end of the lactation cycle when the cow was being milked (d); n is the number of weeks in a single lactation (3)
$MY_{peak} = \max(MY_i)_{i=1}^{i=n}$	MY at lactation peak (kg·d <sup>-1</sup> ), as maximum weekly MY in a lactation cycle of n weeks (4)
$T_{peak} = i_{MY_{peak}}$	Time of peak (wk) is the week in the lactation cycle when the MY peak occurs (5)
$int_{dry} = Int_{calving \text{ to calving}} - DIM$	Time interval (d) during a lactation cycle of calving j-1 and before calving j when the cow was no longer being milked (6)
$cumulativeMY_{305d} = \sum_{i=1}^{i=44} MY_i$	Cumulative MY (kg) during a reference lactation cycle of 305 d (7)
$cumulativeMY_{peak} = \sum_{i=1}^{i=T_{peak}} MY_i$	Cumulative MY (kg) up to the lactation peak at T <sub>peak</sub> (8)
$cumulativeMY_{4wk} = \sum_{i=1}^{i=4} MY_i$	Cumulative MY (kg) up to the 4 <sup>th</sup> week of 1 <sup>st</sup> lactation (9)
$cumulativeMY_{24wk} = \sum_{i=1}^{i=24} MY_i$	Cumulative MY (kg) up to the 24 <sup>th</sup> week of 1 <sup>st</sup> lactation (10)
$Persistence =  (MY_{peak} - MY_{36th \text{ wk}})/(T_{peak} - 36) $	Persistence of the lactation curve (kg·d <sup>-1</sup> ), considering as a reference the 36 <sup>th</sup> week of the lactation cycle (11)
$MY \text{ rise rate} = (MY_{peak} - MY_{1st \text{ wk}})/T_{peak}$	MY rise rate (kg·d <sup>-1</sup> ) is the rate of rise in MY between the 1 <sup>st</sup> week of lactation and the lactation peak (12)
$\overline{MP} = \frac{1}{n} \sum_{i=1}^{i=n} MP_i$	Average milk protein (MP) content for the entire lactation cycle (g·kg <sup>-1</sup> ) (13)
$\overline{MF} = \frac{1}{n} \sum_{i=1}^{i=n} MF_i$	Average milk fat (MF) content for the entire lactation cycle (g·kg <sup>-1</sup> ) (14)
$\%T_{MF/MP < 1.0} = \begin{cases} \text{if } MF_i/MP_i < 1.0, (100/n) \sum_{i=1}^{i=n} i \\ \text{if } MF_i/MP_i \geq 1.0, 0 \end{cases}$	Percentage of time during the total lactation period when the ratio MF/MP was lower than 1.0 (higher risk of acidosis) (15)
$\%T_{MF/MP > 1.4} = \begin{cases} \text{if } MF_i/MP_i > 1.4, (100/n) \sum_{i=1}^{i=n} i \\ \text{if } MF_i/MP_i \leq 1.4, 0 \end{cases}$	Percentage of time during total lactation period when the ratio MF/MP was higher than 1.4 (risk of ketosis) (16)
$BW_{calving} = BW_{i=T_{calving}}$	Average weekly body weight (kg) at calving (17)
$BW_{min} = \min(BW_i)_{i=1}^{i=n}$	Minimum average weekly body weight (kg) achieved during the 1 <sup>st</sup> lactation (18)
$BCS_{min} = \min(BCS_i)_{i=1}^{i=n}$	Minimum average weekly body condition score achieved during the lactation period. Ranging between 0 and 5, with 0.25 interim steps. (19)
$mastitis_{count} = \sum_{i=1}^{i=n} mastitis_{count,i}$	Total count of mastitis that occurred during a lactation cycle of n weeks (20)
$lameness_{count} = \sum_{i=1}^{i=n} lameness_{count,i}$	Total count of lameness that occurred during n weeks after calving (21)
$med.int_{count} = \sum_{i=1}^{i=n} med.int_{count,i}$	Total count of medical interventions that occurred during n weeks after calving (22)
$heat_{count} = \sum_{i=1}^{i=n} heat_{count,i}$	Total count of observed heat events (23)
$(AI + NI)_{count} = \sum_{i=1}^{i=n} AI_i + \sum_{i=1}^{i=n} NI_i$	Total count of artificial and/or natural inseminations per lactation (24)
$int_{calving-1st \text{ heat}} = T_{calving} - T_{1st \text{ heat}}$	Time interval (d) between calving and first heat (25)

$\text{int}_{\text{calving}-1^{\text{st}} \text{insemination}} = T_{\text{calving}} - T_{1^{\text{st}} \text{insemination}}$	Time interval (d) between calving and first insemination	(26)
$\text{DO} = T_{\text{calving}} - T_{\text{conception}}$	Days Open: time interval (d) between calving and conception	(27)
$\text{int}_{1^{\text{st}} \text{heat}-\text{conception}} = T_{1^{\text{st}} \text{heat}} - T_{\text{conception}}$	Time interval (d) between first heat and conception	(28)
$\text{CI} = T_{\text{calving } j} - T_{\text{calving } j-1}$	Calving Interval: interval (d) between two subsequent calvings (j-1 and j)	(29)
$\overline{\text{concentrate}}_{10 \text{ wk}} = (1/10) \sum_{i=1}^{10} (\text{concentrate}_i)$	Average concentrate intake ( $\text{kg}_{\text{DM}} \cdot \text{d}^{-1}$ ) up to 10 weeks after calving; $\text{concentrate}_i$ is the average concentrate intake ( $\text{kg}_{\text{DM}} \cdot \text{d}^{-1}$ ) at week i after calving	(30)
$\overline{\text{forage}}_{10 \text{ wk}} = (1/10) \sum_{i=1}^{10} (\text{forage}_i)$	Average forage intake ( $\text{kg}_{\text{DM}} \cdot \text{d}^{-1}$ ) up to 10 weeks after calving; $\text{forage}_i$ is the average forage intake ( $\text{kg}_{\text{DM}} \cdot \text{d}^{-1}$ ) at week i after calving	(31)
$\text{productive lifespan} = \text{age}_{\text{culling}} - \text{age}_{1^{\text{st}} \text{ calving}}$	Time period between first calving and culling event (d)	(32)

647 \*Raw data were presented on a daily basis, but to cope with missing data for the variables that were not measured daily,  
648 average values per week after calving were calculated. For the variables presented in this table, we still present data on a  
649 daily basis by dividing weekly based data by 7.

650

651 Table 2. Descriptive statistics for age at first calving, milk production, body weight, body  
 652 condition score, fertility and health traits and longevity of primiparous cows. N=185

	Mean $\pm$ SE	Maximum	Minimum
Age at first calving (d)	1094 $\pm$ 3	1002	1212
Duration of 1 <sup>st</sup> lactation (d)	400 $\pm$ 3	324	542
Average milk yield over 1 <sup>st</sup> lactation (kg·d <sup>-1</sup> )	18.5 $\pm$ 0.3	6.0	28.3
Average body weight over 1 <sup>st</sup> lactation (kg)	567 $\pm$ 3	455	696
Body weight at 1 <sup>st</sup> calving (kg)	628 $\pm$ 3	506	759
Average body condition score over 1 <sup>st</sup> lactation	1.11 $\pm$ 0.03	0.90	2.58
%T <sub>MF/MP&lt;1.0</sub> (%)	9.12 $\pm$ 0.07	0	68
%T <sub>MF/MP&gt;1.4</sub> (%)	8.21 $\pm$ 0.04	0	62
Mastitis count	0.4 $\pm$ 0.1	0.0	3.0
Lameness count	0.5 $\pm$ 0.1	0.0	4.0
Number of AI and/or NI per conception	1.8 $\pm$ 0.1	1.0	5.0
Interval between calving to conception, or 'days open' (d)	113 $\pm$ 3	39	252
Number of lactations per cow	4.1 $\pm$ 0.1	2	11
Lifespan (d)	2640 $\pm$ 50	1587	4996

653 Data represent mean  $\pm$  SE (standard error of the mean) over the first lactation cycle, except for variables "Number of lactations" and  
 654 "Lifespan"; %T<sub>MF/MP<1.0</sub> and %T<sub>MF/MP>1.4</sub> represent the percentages of time when the ratio MF/MP (milk fat content divided by milk  
 655 protein content) is less than 1.0 and higher than 1.4, respectively, as indicators of acidosis and ketosis. AI and NI stand for artificial and  
 656 natural inseminations, respectively.

657

658



659 Table 3. Summary of results from PCA for the variables that remained in the analysis.  
 660 Eigenvectors of each variable, and correlations between variables with the first four principal  
 661 components (PC)

Variables	Significant correlations (95% probability) between variables and components				VIF
	PC1	PC2	PC3	PC4	
$\overline{MY}_{305d}$ (kg·d <sup>-1</sup> )	<b>0.69</b>	-0.53	-0.16	-	8.2
Int <sub>dry</sub> (d)	<b>0.54</b>	0.47	0.17	-0.24	2.3
cumulativeMY <sub>peak</sub> (kg)	<b>0.77</b>	-0.23	-	-	4.0
cumulativeMY <sub>4wk</sub> (kg)	<b>0.86</b>	-0.24	-	-	8.0
Persistency (kg·d <sup>-1</sup> )	<b>-0.61</b>	0.32	-	0.38	3.6
$\overline{MP}$ (g·kg <sup>-1</sup> )	-0.31	0.26	<b>0.54</b>	-	6.2
BW <sub>calving</sub> (kg)	0.32	-0.40	<b>0.73</b>	-	4.1
BW <sub>min</sub> (kg)	0.21	-0.42	<b>0.77</b>	-	3.6
BCS <sub>min</sub>	-0.32	-0.21	<b>0.59</b>	-	1.6
heat <sub>count</sub>	0.36	<b>0.69</b>	-	-	2.4
(AI + NI) <sub>count</sub>	0.44	<b>0.75</b>	-	-	6.1
int <sub>calving-1<sup>st</sup> heat</sub> (d)	0.53	0.18	-	<b>0.68</b>	2.3
DO (d)	0.54	<b>0.70</b>	0.17	0.23	8.2
$\overline{\text{concentrate}}_{10\text{ wk}}$ (kgDM·d <sup>-1</sup> )	<b>0.75</b>	-0.29	-	-	3.7
productive lifespan (d)	-0.30	<b>-0.34</b>	-	-	1.3
Eigenvalue	4.4	2.9	1.9	1.0	-
Cumulative variability (%)	29	48	61	68	-

662 Bold numbers represent the highest significant correlations between variables and components.  $\overline{MY}_{305d}$  is the average daily milk yield for  
 663 the entire lactation cycle; Int<sub>dry</sub> is the time interval during a lactation cycle of calving j-1 and before calving j when the cow was no longer  
 664 being milked; cumulativeMY<sub>peak</sub> is the cumulative milk yield up to the lactation peak; cumulativeMY<sub>4wk</sub> is the cumulative milk yield up to  
 665 the 4<sup>th</sup> week of lactation; Persistency is the persistency of the lactation curve, considering as a reference the 36<sup>th</sup> week of the lactation cycle ;  
 666  $\overline{MP}$  is the average milk protein content over the entire lactation cycle; BW<sub>calving</sub> is the average weekly body weight at calving; BW<sub>min</sub> is the  
 667 minimum average weekly body weight over the lactation; BCS<sub>min</sub> is the minimum average weekly body condition score achieved during the  
 668 lactation period; heat<sub>count</sub> is the total count of observed heat events; (AI + NI)<sub>count</sub> is the total count of artificial and/or natural  
 669 inseminations per lactation; int<sub>calving-1<sup>st</sup> heat</sub> is the interval between calving and first heat; DO stands for days open;  $\overline{\text{concentrate}}_{10\text{ wk}}$  is  
 670 the average concentrate intake up to 10 weeks after calving.

671  
672

Table 4. Deviations of cluster means from entire experimental herd mean and post-hoc results for between cluster comparisons

Variable	Herd average	Significant differences from population mean		
		Cluster 1	Cluster 2	Cluster 3
Sample size	185	80	28	77
$\overline{MY}$ (kg·d <sup>-1</sup> )	18	-3*** a	ns <sup>b</sup>	3***c
$\overline{MY}_{305d}$ (kg·d <sup>-1</sup> )	18	-2***b	-2***b	3***a
DIM (d)	320	-17***c	50***a	ns <sup>b</sup>
$MY_{peak}$ (kg·d <sup>-1</sup> )	26	-4***c	ns <sup>b</sup>	5***a
$T_{peak}$ (d)	8	-1* <sup>b</sup>	ns <sup>a</sup>	ns <sup>a</sup>
Int <sub>dry</sub> (d)	78	-9***c	22***a	ns <sup>b</sup>
cumulativeMY <sub>305 d</sub> (kg)	5755	-962***c	ns <sup>b</sup>	995***a
cumulativeMY <sub>peak</sub> (kg)	942	-228***c	ns <sup>b</sup>	234***a
cumulativeMY <sub>4wk</sub> (kg)	580	-79***c	ns <sup>b</sup>	84***a
cumulativeMY <sub>24wk</sub> (kg)	3640	-601***c	ns <sup>b</sup>	659***a
Persistency (kg·d <sup>-1</sup> )	-0.0573	0.0120***a	0.0078 <sup>1a</sup>	-0.0153***b
MY rise rate (kg·d <sup>-1</sup> )	11	ns <sup>a</sup>	ns <sup>a</sup>	ns <sup>a</sup>
$\overline{MP}$ (g·kg <sup>-1</sup> )	38	ns <sup>a</sup>	1 <sup>1a</sup>	ns <sup>a</sup>
$\overline{MF}$ (g·kg <sup>-1</sup> )	33	1 <sup>1a</sup>	1* <sup>a</sup>	-1* <sup>a</sup>
%T <sub>MF/MP&lt;1.0</sub> (%)	8	ns <sup>a</sup>	ns <sup>a</sup>	-1* <sup>b</sup>
%T <sub>MF/MP&gt;1.4</sub> (%)	8	-2***b	ns <sup>a</sup>	ns <sup>a</sup>
BW <sub>calving</sub> (kg)	628	-12 <sup>b</sup>	ns <sup>b</sup>	16* <sup>a</sup>
BW <sub>min</sub> (kg)	564	ns <sup>b</sup>	ns <sup>b</sup>	11 <sup>1a</sup>
BCS <sub>min</sub>	1.2	0.1* <sup>a</sup>	ns <sup>a</sup>	ns <sup>b</sup>
mastitis <sub>count</sub>	0.4	ns <sup>a</sup>	ns <sup>a</sup>	0.1 <sup>1a</sup>
lameness <sub>count</sub>	0.5	-0.2 <sup>b</sup>	ns <sup>a</sup>	0.2 <sup>1a</sup>
med.int <sub>count</sub>	1.5	ns <sup>a</sup>	ns <sup>a</sup>	ns <sup>a</sup>
heat <sub>count</sub>	1.8	-0.4***c	1.4***a	ns <sup>b</sup>
(AI + NI) <sub>count</sub>	1.8	-0.5***c	1.7***a	ns <sup>b</sup>
int <sub>calving-1<sup>st</sup> heat</sub> (d)	72	-17***b	20* <sup>a</sup>	10* <sup>a</sup>
int <sub>calving-1<sup>st</sup> insemination</sub> (d)	88	-8* <sup>b</sup>	ns <sup>a</sup>	ns <sup>a</sup>
DO (d)	113	-26***c	67***a	ns <sup>b</sup>
int <sub>1<sup>st</sup> heat-conception</sub> (d)	41	-10* <sup>b</sup>	48***a	-7 <sup>b</sup>
CI (d)	398	-26***c	72***a	ns <sup>b</sup>
forage <sub>10 wk</sub> (kgDM·d <sup>-1</sup> )	9.6	ns <sup>a</sup>	ns <sup>a</sup>	ns <sup>a</sup>
concentrate <sub>10 wk</sub> (kgDM·d <sup>-1</sup> )	5.7	-0.9***c	ns <sup>b</sup>	1.0***a
productive lifespan (d)	1546	167* <sup>a</sup>	-424* <sup>c</sup>	ns <sup>b</sup>

673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694

$\overline{MY}$  is the average daily milk yield for the entire lactation cycle;  $\overline{MY}_{305d}$  is the average daily milk yield for up to a standard lactation cycle of 305 days; DIM stands for days in milk;  $MY_{peak}$  is the milk yield corresponding to the time when the lactation cycle reaches its peak;  $T_{peak}$  is the time in the lactation cycle when the lactation peak occurs; Int<sub>dry</sub> is the time interval during a lactation cycle of calving j-1 and before calving j when the cow was no longer being milked; cumulativeMY<sub>305 d</sub> is the cumulative milk yield during a reference lactation cycle of 305 days ; cumulativeMY<sub>peak</sub> is the cumulative milk yield up to the lactation peak; cumulativeMY<sub>4wk</sub> is the cumulative milk yield up to the 4<sup>th</sup> week of lactation; cumulativeMY<sub>24wk</sub> is the cumulative milk yield up to the 24<sup>th</sup> week of lactation; Persistency is the persistency of the lactation curve, considering as a reference the 36<sup>th</sup> week of the lactation cycle; MY rise rate is rate of rise in daily milk yield between the 1<sup>st</sup> week of lactation and the lactation peak;  $\overline{MP}$  is the average milk protein content over the entire lactation cycle;  $\overline{MF}$  is the average milk fat content over the entire lactation cycle; %T<sub>MF/MP<1.0</sub> and %T<sub>MF/MP>1.4</sub> represent the percentages of time when the ratio MF/MP (milk fat content divided by milk protein content) is less than 1.0 and higher than 1.4, respectively, as indicators of acidosis and ketosis; BW<sub>calving</sub> is the average weekly body weight at calving; BW<sub>min</sub> is the minimum average weekly body weight over the lactation; BCS<sub>min</sub> is the minimum average weekly body condition score achieved during the lactation period; mastitis<sub>count</sub> and lameness<sub>count</sub> are the total counts of mastitis and lameness, respectively, that occurred during the lactation cycle; med.int<sub>count</sub> is the total count of medical interventions that occurred during the lactation cycle; heat<sub>count</sub> is the total count of observed heat events; (AI + NI)<sub>count</sub> is the total count of artificial and/or natural inseminations per lactation; int<sub>calving-1<sup>st</sup> heat</sub> is the interval between calving and first heat; int<sub>calving-1<sup>st</sup> insemination</sub> is the time interval between calving and first insemination; DO stands for days open; int<sub>1<sup>st</sup> heat-conception</sub> stands for time interval between first heat and conception; CI stands for calving interval; forage<sub>10 wk</sub> is the average forage intake up to 10 weeks after calving; concentrate<sub>10 wk</sub> is the average concentrate intake up to 10 weeks after calving; Differences followed by “\*”, “\*\*”, “\*\*\*” and “\*\*\*\*” have P-values of <0.05, <0.01, <0.001 and <0.0001; “ns” means that cluster mean is not significantly different from the entire herd mean; Letters stem from results of post-hoc tests: Tukey and Wilcoxon signed rank tests for normally and non-normally distributed variables, respectively (statistical significance at P≤0.05). Variables presenting same letters per cluster denote that cluster means are not significantly different.

695

Table 5. Count of cows per cluster and reason for culling

Cluster	1	2	3
Reason for culling			
Poor fertility	39	21	49
Poor health	26	7	16
Other reasons	15	0	12

696

697

698

699

Reasons for culling were: Poor fertility (including empty, abortion, late calving and calving troubles); poor health (including mastitis, lameness, acidosis, other infections and overall health status); (3) poor production performance, (4) herd renewal (including old age), and (5) accidents on farm. Results from Fisher's exact test of contingents led us to reject the null hypothesis of independence between counts of cows per cluster and reason for culling ( $P = 0.023$ ).

700

701 Table 6. Quality assessment of calibrated and cross-validated (between brackets) linear  
 702 discriminant models

Cluster	Predicted classification (count)*			Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
	1	2	3				
1	77(70)	<b>0 (0)</b>	<b>3 (10)</b>	96 (94)	98 (89)	98 (96)	
2	<b>1 (3)</b>	27 (21)	<b>0 (4)</b>	96 (86)	100 (97)	100 (97)	97 (89**)
3	<b>1 (6)</b>	<b>0 (4)</b>	76 (67)	96 (86)	97 (96)	96 (89)	

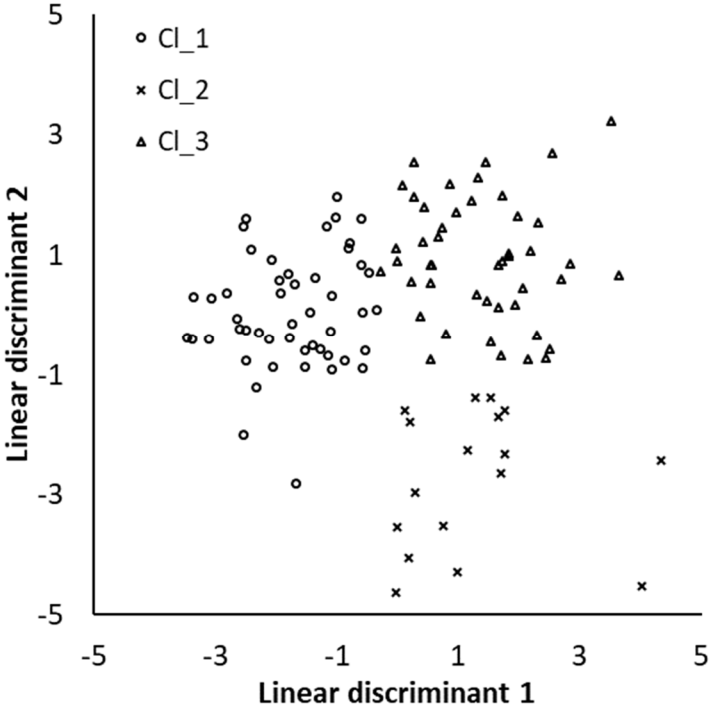
703 \*Values in **bold** represent the percentage of individuals from a given cluster which were misclassified in other clusters by the linear  
 704 discriminant method; \*\*Accuracy value obtained from cross-validation is significantly different from zero at a confidence of 95%.

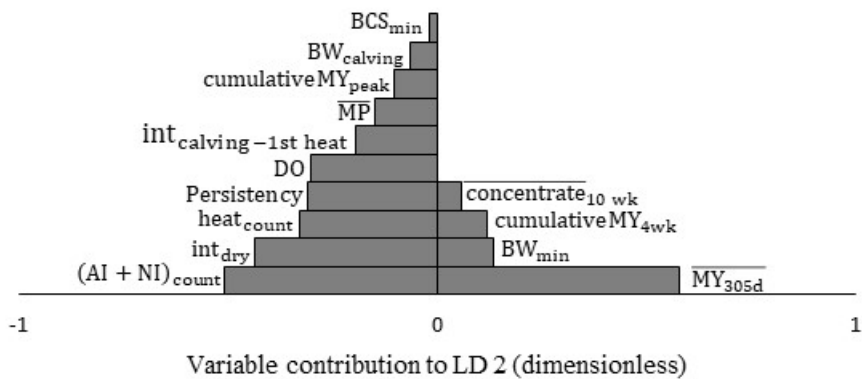
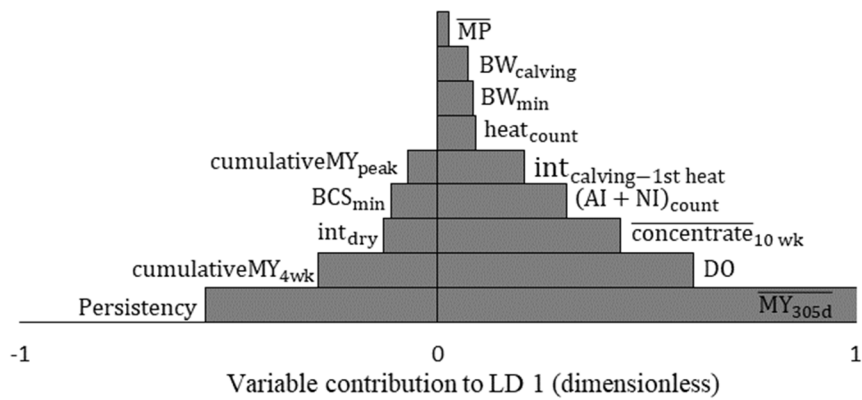
705

706

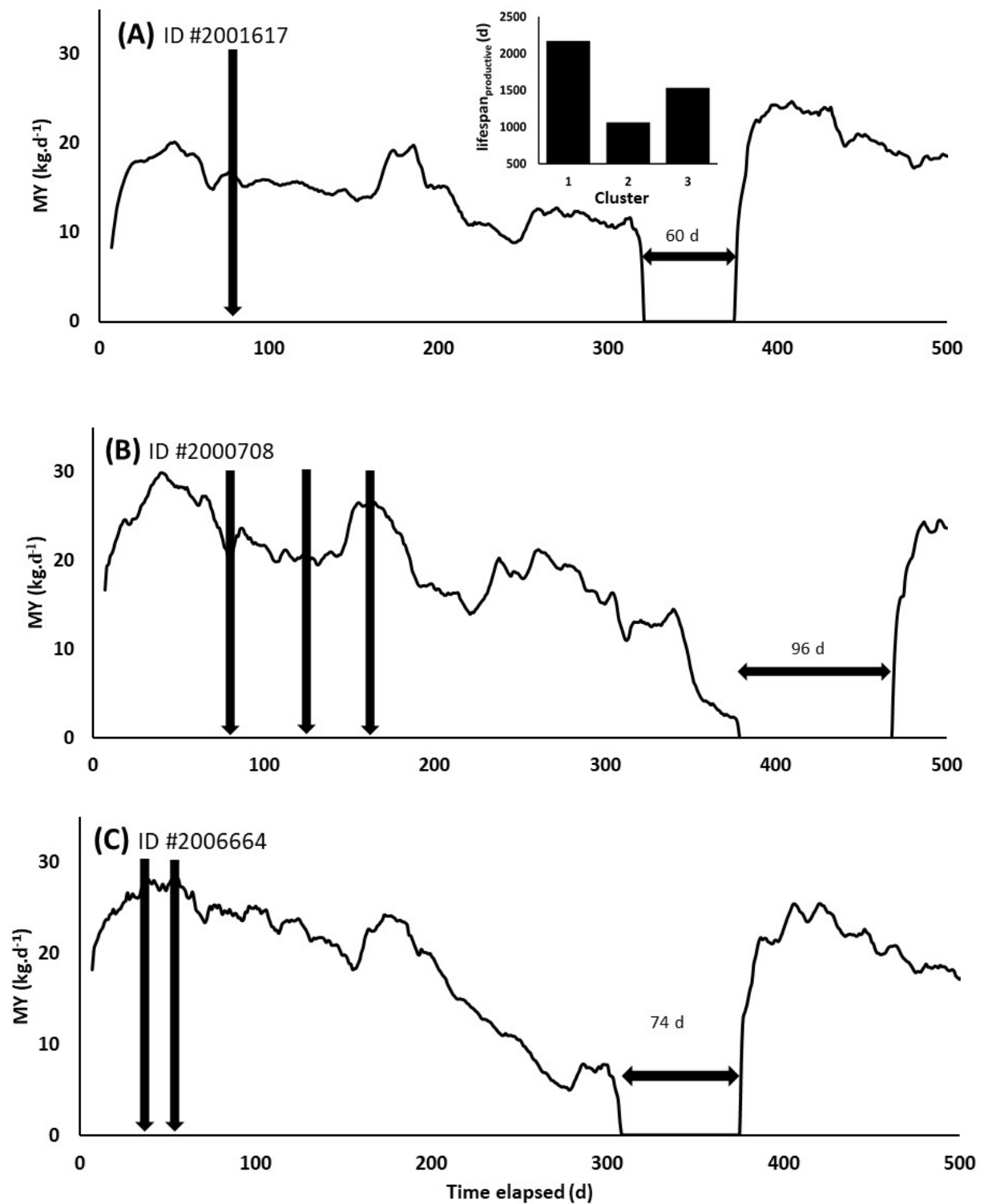
707

Mendes1





$\overline{MY}_{305d}$  is the average daily milk yield for up to a standard lactation cycle of 305 days;  $int_{dry}$  is the time interval during a lactation cycle of calving  $j-1$  and before calving  $j$  when the cow was no longer being milked;  $cumulativeMY_{peak}$  is the cumulative milk yield up to the lactation peak;  $cumulativeMY_{4wk}$  is the cumulative milk yield up to the 4<sup>th</sup> week of lactation; Persistenceency is the persistency of the lactation curve, considering as a reference the 36<sup>th</sup> week of the lactation cycle;  $\overline{MP}$  is the average milk protein content over the entire lactation cycle;  $BW_{calving}$  is the average weekly body weight at calving;  $BW_{min}$  is the minimum average weekly body weight over the lactation;  $BCS_{min}$  is the minimum average weekly body condition score achieved during the lactation period;  $heat_{count}$  is the total count of observed heat events;  $(AI + NI)_{count}$  is the total count of artificial and/or natural inseminations per lactation;  $int_{calving-1st\ heat}$  is the interval between calving and first heat; DO stands for days open  $\overline{concentrate}_{10\ wk}$  is the average concentrate intake up to 10 weeks after calving.



A paragon is a representative cow selected per cluster based on two criteria: (1) the shortest geometrical distance (calculated with individual coordinates obtained from principal component analysis) of the individual to the cluster barycenter, and (2) the individual representativeness of cluster characteristics, i.e. whether or not paragons resemble average individuals per cluster.