Supplementary material

Modeling cow somatic cell count using sensor data as input to generalized additive models

Dorota Anglart[1, 2,*], Charlotte Hallén-Sandgren[1], Patrik Waldmann[3], Martin Wiedemann[4] and Ulf Emanuelson[2]

[1]DeLaval International AB, PO Box 39, SE-147 21, Tumba, Sweden

[2]Swedish University of Agricultural Sciences, Department of Clinical Sciences, PO Box 7054, SE-750 07 Uppsala, Sweden

[3]Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, PO Box 7023, SE-750 07 Uppsala, Sweden

[4]DeLaval GmbH, Wilhelm-Bergner-Str. 5, 21509 Glinde, Germany

S1 Materials and Methods *Data collection and Data preparation*

Supplementary Table S1: Data retrieved from the automatic milking rotary system at the quarter and cow levels, cleaning principles for each item (<1% data removed). New variables were created for conductivity, mean flow and peak flow and according to a, b, c, d, e and f (found in footnote). The new variables for kick off, incomplete milking and not milked were created according to g, h and i, Blood according to h and j while new variable for milk yield was created according to k.

Item	Explanation (units)	Removed observation
Quarter level		
Conductivity[a,b,c,d,e,f]	Millisiemens (mS/cm)	If conductivity < 3 or > 10
Mean flow[a,b,c,d,e,f]	Average milk flow (g/min)	If flow ≤ 180 or ≥ 1800
Peak flow[a,b,c,d,e,f]	Maximum milk flow, 5 seconds rolling average (g/min)	If flow ≤ 180 or ≥ 2600
Incomplete[a,b,c,d,e,f]	"1" when milk yield <50% of expected yield, under the condition:expected yield >1 kg, milk yield >3 kg (1/0)	If quarter set not to be milked
Kick off[g,h,i]	"1" if one or several cups were detached during milking (1/0)	If quarter set not to be milked
Not milked[g,h,i]	Milk yield <0.2 kg (1/0)	If quarter set not to be milked
Blood[h,j]	Red color of milk, parts per million (ppm)	If quarter set not to be milked
Milk yield[k]	Milk yield measured by milk meter (kg)	If quarter set not to be milked
Expected milk yield	Based on previous milk yield and milking interval (kg)	If quarter set not to be milked
Cow level		
MDi	Mastitis detection index based on different phases of conductivity and blood in milk (unitless)	
Udder counter	If MDi >1.4 counting up, if MDi <1.4 counting down for last 10 milkings (1-10)	
Milking unit	(1–24)	
Milking duration	Total milking time from attachment of first cup to removal of last (minutes)	Time ≤ 3 or ≥ 17

[a].flag.diff.milking = dichotomized, scored "1" if change from previous corresponding milking session was larger or smaller than $2 \times SD$ (e.g., Conductivity.flag.diff.milking)

[b].diff = highest value minus lowest value within cow and milking session

[c].max = highest value (quarter) within cow and milking session

[d].mean = arithmetic mean all quarters within cow and milking session

[e].min = lowest value (quarter) within cow and milking session

[f].var = variance between quarters within cow and milking session

[g].quarter = dichotomized, "1" if the event occurred for a quarter (e.g., Incomplete.quarter)

[h].score = number (0–4) of quarters involved in the event (i.e., Incomplete.score)

[i].7 = 7-day rolling average combined with the "score" variable (e.g., Incomplete.score.7)

[j].1000 or .2000 = two variables: a threshold of ppm ≤ 1000 or ppm ≤ 2000 was added to the dichotomization: "1" if there was occurrence of blood above the thresholds (e.g., Blood.score.1000, Blood.score2000)

[k].diff.milkings = deviation from previous corresponding milking session at the quarter level

S2 Results *Descriptive Statistics*

The complete and edited milking dataset contained 2384 observations with possible explanatory variables from 372 cows, of which 319 had complete records with 1758 observations from all 8 CMSCC sampling events. The distribution of cows among LN groups 1, 2, and ≥3 was 119, 90, and 110, respectively, and the average DIM on the test day was 191. The average CMSCC was 68,000 cells/mL and the interquartile range was 26,000–125,000 cells/mL.

Supplementary Table S5: Descriptive statistics per milking for the composite milk somatic cell count (CMSCC) and data received from the automatic milking rotary system

Numerical variable[a]	Min	First quantile	Median	Third quantile	Max	Observations (n)
CMSCC ($\times 10^3$ cells/mL)	0	26	53	125	6766	2384
Days in milk	8	116	199	263	631	2382
Milking time (min)	3	5	7	8	17	2387
Milk yield (kg)	4.8	14.2	16.7	19.2	28.7	2227
MDi[b]	0.98	1.14	1.22	1.34	4.12	2376
Mean flow[c] (g/min)	0.19	0.65	0.81	0.96	1.76	9245
Peak flow[c] (g/min)	0.38	1.17	1.38	1.62	2.58	9292
Conductivity[c] (mS/cm)	4.50	5.75	6.05	6.41	9.15	9283

Categorical item[d]	Number of levels	Level 0 (n)	Level 1 (n)
Incomplete[c]	2	9359	32
Kick off[c]	2	8683	712
Not milked[c]	2	9356	33
Udder counter	11	2301	83[e]

[a] Treated as continuous numeric variables in the analysis; [b] Mastitis detection index; [c] Quarter variables; [d] Treated as factor variables in the analysis with *n* = unique number of levels; [e] *n* level ≥ 1 presented