Computers and Electronics in Agriculture 109 (2014) 191–194

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Application note

Dynamic monitoring of reproduction records for dairy cattle

C. Cornou [a,*], S. Østergaard [b], M.L. Ancker [c], J. Nielsen [c], A.R. Kristensen [a]

[a] HERD – Centre for Herd-oriented Education, Research and Development, Department of Large Animal Sciences, University of Copenhagen, Grønnegårdsvej 2, 1870 Frederiksberg C, Denmark
[b] Department of Animal Science, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark
[c] Knowledge Center for Agriculture, Agro Food Park 15, Skejby, 8200 Aarhus N, Denmark

ARTICLE INFO

Article history:
Received 2 May 2014
Received in revised form 27 August 2014
Accepted 27 September 2014

Keywords:
Dairy cow
Dynamic generalized linear model
Monitoring system
Reproduction result

ABSTRACT

This application note presents a newly developed surveillance module for monitoring reproduction performances in dairy herds. It is called Critical Control Point and is part of a recently developed management tool, Dairy Management System. This management tool is commercialized as software intended both for farmers, extension officers, breeding advisors and veterinarians. Insemination and conception rates, for cows and heifers, are modeled at the herd level using Dynamic Generalized Linear Models for binomial data. The results are updated and monitored on a weekly basis, using control charts, and alarms are provided when the performances are below target values. Both the number of observed inseminations and pregnancies, and the insemination and pregnancy rates are monitored. The components of the user interface are presented and some comprehensive graphs, accessible to the user, illustrate the herd's performances over the last 52 weeks.

1. Introduction

Nowadays, large amounts of data for calvings, inseminations and pregnancy tests are systematically collected on dairy farms. A common form for monitoring data in livestock production systems is the use of statistical control charts (de Vries and Reneau, 2010), where raw or pre-processed data are plotted over time in order to control whether a process is running as it should (Montgomery, 2005).

In Denmark, Thysen and Enevoldsen (1994) implemented a monitoring system that led to the commercialization of a software for monitoring reproduction results: HerdView. Like all other dairy herd management information systems in Denmark, HerdView relies on extraction of raw data from a national database run by Dairy and Cattle Farming, Knowledge Centre for Agriculture. In the case of HerdView, data is not extracted online, but manually. The standard tool for dairy farmers for online access to the database is a separate software system, called Animal Registration. So far, the reproduction management system provided through Animal Registration uses a set of monthly and yearly key figures basically consisting of: start of insemination (days after calving), insemination rate and conception rate. These are reported (i) as percentage for the last 12 months and (ii) as monthly averages, in order to visualize the evolution of the key figures.

The current Animal Registration system will in the future be replaced and extended by a new Dairy Management System (DMS) which is developed and provided by the Danish Knowledge Centre for Agriculture. This management tool can be used by farmers as well as by extension officers, breeding advisors and veterinarians and it is gradually being extended. The DMS holds different modules with different purposes including: (1) forecasting of production results for the next few years, (2) budgeting of feeding, (3) ration formulation, (4) economical evaluation of production results. During 2012, DMS was extended with a surveillance module which holds a number of Critical Control Points (CCPs) concerning milk production and milk quality, reproduction, animal health, energy utilization and economy.

A recently developed and implemented part of the CCP allowing a dynamic monitoring of reproduction results, on a weekly basis, is presented in this paper. It consists of a dynamic monitoring of inseminations and pregnancies (both their number and rate), on a weekly basis. As opposed to HerdView, this system is fully integrated with the national database.

2. Model and methods

2.1. Data monitored

The data consist of weekly counts of inseminations, eligible cows for inseminations, and pregnancies. The period of eligibility stretches from the end of the voluntary waiting period (after

* Corresponding author.
E-mail address: cec@sund.ku.dk (C. Cornou).

http://dx.doi.org/10.1016/j.compag.2014.09.017
0168-1699/© 2014 Elsevier B.V. All rights reserved.

calving) until a successful insemination or a decision to stop inseminations has taken place. Results of these variables are usually computed as success rates:

- The weekly insemination rate is computed using the number of inseminated cows and the number of cows eligible for breeding at week t;
- The weekly conception rate is computed from the number of pregnant cows obtained from pregnancy checks by rectal palpation and the number of inseminated cows at week t.

In the following, κ_t will refer to the observed inseminations (or pregnancies) at week t, N_t to the eligible cows for breeding (or tested for pregnancy) at week t, and p_t to the insemination rate/pregnancy rate at week t, respectively.

2.2. Dynamic generalized linear models

The outcome of inseminations or pregnancy tests follows a Binomial distribution, defined as the distribution of successes κ_t for N_t independent observations having a probability p_t of success. West and Harrison (1997) described Dynamic Generalized Linear Models (DGLMs) as a tool for dynamic monitoring of binomial data. A DGLM for binomial data with the methodological modifications developed by Bono et al. (2013) was therefore implemented, using the natural parameter η_t, defined as

$$\eta_t = \text{logit}(p_t), \tag{1}$$

where p_t is the underlying probability of success (or success rate) at time t and logit is the logistic transform.

The observation equation of the DGLM is defined as

$$(\kappa_t | \eta_t, N_t) \sim \mathcal{B}(N_t, p_t), \tag{2}$$

with κ_t following a binomial distribution with parameters N_t and $p_t = \text{expit}(\eta_t)$, where expit is the inverse logistic transform. The evolution over time of the natural parameter η_t is defined in the system equation as

$$\eta_t = \eta_{t-1} + \omega_t, \quad \omega_t \sim [0, W], \tag{3}$$

where the system errors ω_t are assumed independent and identically distributed with mean 0 and constant variance W. Then, for each week:

- Forward filtering of the data is performed using the updating equations of the DGLM, as described by Bono et al. (2013). The filtering provides estimates of the mean (m_t) and variance (C_t) of the underlying value η_t. The probability parameter corresponding to the mean is $p_t = \text{expit}(m_t)$. This filtering process reduces some random binomial variation.
- Backwards smoothing of the data provides the smoothed mean $\tilde{m}_t$ and variance $\widetilde{C}_t$ of the underlying value η_t. The smoothing procedure uses all data available at time t. It reexamines the data starting from t and 52 weeks backwards. This retrospective analysis further reduces the random binomial variation, and enhances the understanding of some given events that may have influenced the results.

A more detailed description of the updating equations of the DGLM, for filtering and smoothing, is found in Bono et al. (2013) and a comprehensive application of DGLM for univariate binomial data is found in Cornou and Lundbye-Christensen (2012).

2.3. Estimation of variance components

The constant system variance W (Eq. (3)) is optimized by minimizing the sum of the squares (SS) of the model's forecast errors, e_t, with values for W ranging from 0.001 to 0.1 by step of 0.001.

$$SS = \sum_{t=1}^{T} e_t^2, \quad \text{where } e_t = \kappa_t - \mu_t, \tag{4}$$

with μ_t being the model's forecast mean for observed data. According to Bono et al. (2013), the mean (μ_t) and variance (Σ_t) of the forecast are

$$\mu_t = N_t \text{expit}(m_t), \tag{5}$$

$$\Sigma_t = \Delta_t + \Delta_t^2 (C_{t-1} + W), \tag{6}$$

where N_t is the total number of observations (i.e. eligible/tested) at time t, and $\Delta_t = N_t \text{expit}(m_t)(1 - \text{expit}(m_t))$.

The results of the DGLM are monitored using control charts inspired by Shewhart (Montgomery, 2005). A control chart is composed of three elements: a central line (CL), corresponding to a target value; an upper control limit (UCL) and a lower control limit (LCL).

A first monitoring method consists of a weekly control of the number of observed events κ_t (observed inseminations or pregnancies) compared to the CL, set as the model's forecast μ_t (Eq. (5)). The control limits are drawn using a 95% confidence interval based on the forecast variance (Eq. (6)). An alarm is triggered when κ_t is below the LCL. No alarm is triggered when κ_t exceeds the UCL. Fig. 1(a) illustrates the application of a control chart for monitoring insemination numbers for one herd, for a period of 12 concluded weeks and the current week. An alarm is triggered on week 40. The data mark in week 46 is updated on a daily basis and indicates the number of inseminations performed, so far, in the current week.

The second monitoring method consists of a weekly control of the estimated probability $\tilde{p}_t$ corresponding to the smoothed mean $\tilde{m}_t$ of the underlying value η_t, i.e.

$$\tilde{p}_t = \text{expit}(\tilde{m}_t) \tag{7}$$

compared to the target value p'_t (CL) defined by the farmer, i.e. the expected insemination or conception rate. The upper and lower control limits, UCL and LCL, delimit an approximate 95% confidence interval, and are computed using the smoothed variance $\widetilde{C}_t$. An alarm is triggered when p'_t is above UCL. No alarm is triggered when p'_t is below LCL.

3. Results and implementation

3.1. Variance estimation

Estimation of the system variance was performed for cows and heifers, both for insemination and conception rates, for three breeds: Holstein, Red Danish and Jersey, issued from 218 Danish herds. In practice, due to the difficulty of identifying which breed should be considered, especially in herds with several breeds, it was decided to use the system variance estimated from all breeds. Results of variance estimation are shown is Table 1. This approach is considered satisfying since no crucial difference was observed between the estimates of the individual breeds.

3.2. Implementation

Every morning, a procedure in the central database counts the key values κ_t and N_t for each herd and for four data groups: cows and heifers, and insemination and conception. The number of cows