Animal (2014), **8:3**, pp 477–483 © The Animal Consortium 2014
doi:10.1017/S1751731113002371

Evaluating somatic cell scores with a Bayesian Gaussian linear state-space model

J. Detilleux[1†], L. Theron[2], E. Reding[3], C. Bertozzi[3] and C. Hanzen[2]

[1]*Department of Animal Production, Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Belgium;* [2]*Large Animal Clinic, Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Belgium;* [3]*Association Wallonne de l'Elevage, 4 rue de Champs Elysées, 5590 Ciney, Belgium*

(Received 10 December 2012; Accepted 30 November 2013; First published online 6 January 2014)

Because accurate characterization of health state is important for managing dairy herds, we propose to validate the use of a linear state-space model (LSSM) for evaluating monthly somatic cell scores (SCSs). To do so, we retrieved SCS from a dairy database and collected reports on clinical mastitis collected in 20 farms, during the period from January 2008 to December 2011 in the Walloon region of Belgium. The dependent variable was the SCS, and the independent variables were the number of days from calving, year of calving and parity. The LSSM also incorporated an error-free underlying variable that described the trend across time as a function of previous clinical and subclinical status. We computed the mean sum of squared differences between observed SCS and median values of the posterior SCS distribution and constructed the receiver operating characteristic (ROC) curve for SCS thresholds going from 0 to 6. Our results show SCS estimates are close to observed SCS and area under the ROC curve is higher than 90%. We discuss the meaning of the parameters in light of our current knowledge of the disease and propose methods to incorporate, in LSSM, this knowledge often expressed in the form of ordinary differential equations.

Keywords: bovine mastitis, linear state-space model, somatic cell count

Implications

To decrease mastitis frequency, it is necessary to have adequate measures of the biological mechanisms underlying the disease. Here, we applied a simple linear state-space model on monthly milk somatic cell scores collected routinely in dairy databases. We show the model is highly accurate, very flexible and relatively easy to implement. It allows the identification of disease sub-phenotypes corrected for measurement errors and accounts for the dynamics of the response to the infection. As such, it may shorten the gap between genotype and phenotype expression and help design personalized treatment.

Introduction

Bovine mastitis is a frequently occurring disease with great economic consequences. To find tools that will decrease its frequency, it is necessary to have adequate measures of the biological mechanisms underlying the disease, which is not easy to obtain in field conditions. Here, we propose to combine clinical and subclinical information (obtained in a field study) in a state-space model to obtain measures representative of the defence mechanisms used by the mammary gland to fight infection.

Information on clinical cases is valuable because cows that experienced a case previously have a greater probability of developing a subsequent one (Steeneveld *et al.*, 2008). In addition, records on clinical cases provide continuous information not limited to recording at fixed intervals. However, in practice, data collection on clinical cases depends heavily on the willingness of the observers to collect and report the information so that data may be sparse and subjective.

Information on subclinical cases is also important because defence mechanisms operate before the apparition of clinical signs. To detect subclinical mastitis, owners of automatic milking systems have access to in-line milk-sensing measures such as milk electrical conductivity, yield or temperature, but the frequent occurrence of false-positive alerts has thwarted their widespread use (Hovinen and Pyorala, 2011). Researchers have also looked for biomarkers such as milk lactoferrin, haptoglobin or NGase (Soyeurt *et al.*, 2012), but their measurements are often costly on a routine basis. The most frequent method used to detect subclinical mastitis is based on the measurement of patterns of monthly milk somatic cell counts (SCC). Indeed, SCCs are typically elevated in the presence of intramammary infection and are commonly recorded at (usually) monthly milk recording visits.

† E-mail: jdetilleux@ulg.ac.be

Many models have been proposed to analyze SCC patterns. Some are based on logistic regression (Kristula *et al*., 1992) or alternative traits (de Haas *et al*., 2004). Others draw on mixture (Detilleux and Leroy, 2000; Jamrozik and Schaeffer 2010) and hidden Markov models (Detilleux, 2011; Robertson *et al*., 2011). Advantages of mixture and hidden Markov models include their flexibility associated with their ability to represent observed and hidden states, and their capacity to cope with missing observations. However, they are limited by the finite number of possible hidden states (e.g. infected or not).

State-space models are based, like hidden Markov models, on the assumption that an unobserved variable explains the observed variation and evolves with Markovian dynamics. The main difference is that hidden Markov models use a discrete hidden state variable with arbitrary dynamics, whereas state-space models use a continuous state variable with parametric forms for the transition dynamics (Roweis and Ghahramani, 1999). In the context of bovine mastitis, hidden variables may represent defence mechanisms against mastitis pathogens that cannot be observed accurately under field conditions. For example, SCCs are a mix of leukocytes and epithelial cells (Bradley and Green, 2005). Of both cell types, leukocytes serve as a major defense mechanism to fight disease infection (even if the role of epithelial cells should not be disregarded). If hidden variables are more accurate as indicators of cell dynamics during infection, they will be useful in the search for personalized treatment or for genes and gene pathways that are altered in response to the presence of pathogenic bacteria (Pighetti and Elliott, 2011).

State-space models are also particularly well suited for Bayesian model-fitting approaches and this permits a great degree of flexibility including, for example, nonlinear, non-Gaussian or multivariate processes. Moreover, time steps do not have to correspond to a fixed unit of real time or to follow equally spaced time intervals.

Therefore, the goal of this paper is to validate the use of a linear state-space model (LSSM) for evaluating monthly somatic cell scores (SCSs).

Material and methods

Data description

Data came from a survey on 20 commercial dairy farms (mean of 81 cows per herd) conducted between January 2008 and January 2012, in the Walloon region of Belgium. Herd size, housing systems, milk production and SCCs are described elsewhere (Detilleux *et al*., 2012). Herds were enrolled in the national dairy herds recording system from which SCC data were obtained. We asked participants to record all clinical mastitis events on a web-based interface (Reding *et al*., 2012). We combined data on SCC and clinical mastitis, and considered SCC recorded within 10 days before or after a reported clinical event was associated with the clinical case. We gathered information on year of calving (YVEL), parity (PAR), days in milk measured from the previous calving (DIM) and number of days between successive events (LAG).

Clinical mastitis was diagnosed if milk from one or more glands was abnormal in color, viscosity or consistency, with or without accompanying heat, pain or redness. The SCC records greater than 150 000 SCC/ml in first parity and 250 000 SCC/ml in later lactations were considered as being from subclinical mastitis (Schepers *et al*., 1997). We created a variable, called 'CM', that combined information on SCC and clinical events: CM = 0 when SCC was below the threshold and no case was reported (no mastitis); CM = 1 if a case was reported regardless of the SCC value (clinical mastitis); CM = 2 if no case was reported and SCC was above the threshold (subclinical mastitis).

For the statistical analyses, we limited records to the first three lactations and the first 300 days in lactation (or less if the lactation is terminated before the 300th day; DIM $\leqslant$ 300) because few animals had extended lactations. We transformed SCC into linear SCS more normally distributed than SCC (Ali and Shook, 1980): $SCS = 3 + \log_2(SCC/100\,000)$, where SCC was the number of somatic cells per milliliter.

Statistical analyses

We developed below a LSSM to analyze SCS (Chen and Brown, 2013). The first equation (called 'measurement equation') in the LSSM defines how a hidden health variable, called 'HID', affected observed SCS. The next four equations (called 'transition equations') in the LSSM are first-order Markov processes that describe the trend across time in the unobserved HID as a function of previous HID and CM. The model is specified by the distributions of SCS and HID as

$$SCS_i^t \sim \text{Normal}\left(\mu_i^t, \sigma_i^2\right)$$

$$HID_i^t \sim \text{Normal}\left(\delta_i^t, 1\right)$$

with

$$\mu_i^t = a_0 + a_1\, DIM_i^t + a_2 \ln\left(DIM_i^t\right) + a_3\, YVEL_i^t + a_4\, \delta_i^t$$

$$\delta_i^t = h_0 HSC_i^{t-1} + h_1\, \delta_i^{t-1}/LAG_i^{t-1} + h_4 PAR_i^t \quad \text{if } CM_i^{t-1} = 0$$

$$\delta_i^t = h_0 HSC_i^{t-1} + h_2 \delta_i^{t-1}/LAG_i^{t-1} + h_4 PAR_i^t \quad \text{if } CM_i^{t-1} = 1$$

$$\delta_i^t = h_0 HSC_i^{t-1} + h_3 \delta_i^{t-1}/LAG_i^{t-1} + h_4 PAR_i^t \quad \text{if } CM_i^{t-1} = 2$$

$$\delta_i^1 = h_x + h_4 PAR_i^1$$

where the index i is for the cow, and t is for the control number within the lactation ($t = 1, 2, \ldots, T_i$). The dependent variables are SCSs and hidden variable (HID). We regarded HID as error-free indicators of cell dynamics

because they are corrected for errors in the measurement equation for SCS. Otherwise, we would have had, for $CM_i^{t-1} = 0, \mu_i^t = a_0 + a_1 DIM_i^t + a_2 \ln(DIM_i^t) + a_3 YVEL_i^t + h_0 HSC_i^{t-1} + h_1 \mu_i^{t-1}/LAG_i^{t-1} + h_4 PAR_i^t$ when we have $\mu_i^t = a_0 + a_1 DIM_i^t + a_2 \ln(DIM_i^t) + a_3 YVEL_i^t + a_4 h_0 HSC_i^{t-1} + a_4 h_1 \delta_i^{t-1}/LAG_i^{t-1} + a_4 h_4 PAR_i^t$. The HIDs are also standardized with a variance set to 1. The herd score (HSC) is the arithmetic mean of individual cow SCS for each milk recording date (Lievaart *et al*., 2007). Days in milk (DIM) and year of calving (YVEL) are fixed effects potentially affecting SCS (Harmon, 1994), whereas parity (PAR) and HSC may act indirectly on SCS through their effects on HID. Although we assumed the same coefficients h_0 and h_4 for all states, and therefore considered previous values for PAR and HSC had the same effects on SCS regardless of the state, we put PAR and HSC on the transition equations because we believe, at least conceptually, they acted more directly on HID than on SCS. Indeed, among others, Laevens *et al.* (1997) were unable to find a significant effect of parity on SCS in bacteriologically negative cows, and Faye *et al.* (1994) showed frequency of clinical mastitis is higher in herds with high bulk tank SCC. Transition between successive HID depends on the HID and CM observed at the previous record (one-period lagged value) weighted by the number of days between successive records (LAG). We assumed error terms to be independent and time invariant.

We chose inverse-gamma priors, IG(0.001, 0.001), for the variance component parameters (σ_i^2) and normal priors for the coefficients a_0 to a_4, h_x, and h_0 to h_4. Initial values for the mean and precision of all normal priors were set at 0 and 10^{-5}, respectively. We implemented the model in Openbugs (Lunn *et al*., 2009) for each herd separately. For two herds (results not shown), we ran multiple chains of 50 000 iterations (5000 burn in and 45 000 sampling iterations) with different starting values, checked the Gelman and Rubin statistics (1996) and visually inspected the trace plots. The Gelman and Rubin statistics were close to 1 and chains had a tendency to converge much earlier than 5000 iterations. This is likely because of the large sample sizes with a large amount of information in the data for the parameters. However, the computing time was long (around 6 h on PC per herd), and therefore we ran unique chains of 50 000 runs for each of the remaining herds.

Comparisons between estimated and observed values

Median $(\hat{SCS}_i^t)$ values of the posterior SCS distribution were compared with the corresponding observed values (SCS_i^t) and the mean sum of squared residuals was calculated for each herd: $MSE_h = \sum (\hat{SCS}_i^t - SCS_i^t)^2/n_h$ for $t=1$ to T_i and $i=1$ to n_h, where n_h is the number of cows recorded in the hth herd. Next, we constructed the receiver operating characteristic (ROC) curves for SCS values above thresholds going from 0 to 6. Thresholds 0 to 6 span a wide range of SCS values observed in healthy and infected cows and correspond to SCC from 0 to 800 000 cells/ml. A ROC curve is the plot of true-positive rates against false-positive rates for the different threshold values. Thus, a model with high discriminating power has a ROC curve close to the upper left-hand corner of the plot (Uhler, 2009). We also computed the area under the ROC curve using the trapezoid rule (Metz *et al*., 1998).

Results and discussion

Description of the data set

Descriptive data are in Table 1. A total of 29 858 records were collected from 2729 lactations. The incidence of reported cases varied from 0 to 20 cases per 100 cow-years at risk, which is lower than some estimates found in the literature. Indeed, Bradley *et al*. (2007) and Miltenburg *et al*. (1996) reported mean incidence of clinical mastitis ranging from 47% to 71% in England and Wales, and from 13% to 30% in the Netherlands, respectively. In five Holstein herds in the Czech Republic, the overall incidence was 0.68 clinical cases per cow per year, varying from 0.35 to 1.45 cases between farms (Wolfova *et al*., 2006). Other estimates were nearer to ours with 14.4 cases per 100 cow-years at risk in Uruguay (Gianneechini *et al*., 2002) and 14.8 in Norway (Osteras *et al*., 2007). Such differences between countries may be associated with factors such as climate, detection and reporting methods, breed, level of production and management. Another possibility is that not all clinical cases

Table 1 *Descriptive statistics for each herd: number of lactations, number of records, mean for SCS and their standard deviation in parentheses, and incidence (%) of reported clinical case per cow-year at risk*

Herds	*n* lactations	*n* records	SCS	Incidence	Herds	*n* lactations	*n* records	SCS	Incidence
1	315	3047	1.53 (1.64)	9.17	11	159	2127	1.75 (1.48)	3.33
2	65	542	2.37 (1.90)	20.10	12	81	1174	2.46 (1.67)	5.74
3	141	1738	2.80 (1.46)	2.45	13	61	733	2.25 (1.55)	0.00
4	116	1523	3.16 (1.59)	2.16	14	86	1281	2.11 (1.81)	1.72
5	96	937	2.45 (1.52)	20.36	15	86	1028	2.35 (1.77)	3.06
6	142	1361	1.73 (1.78)	8.69	16	129	1572	2.25 (1.82)	3.49
7	96	937	1.89 (1.66)	7.19	17	143	1212	1.69 (1.71)	0.00
8	201	1971	2.55 (1.86)	6.64	18	54	613	2.79 (1.60)	20.53
9	113	1141	2.08 (1.79)	3.88	19	221	1877	2.47 (1.95)	15.03
10	253	3067	2.51 (1.77)	1.06	20	171	1977	2.86 (1.63)	18.66

SCS = somatic cell scores.

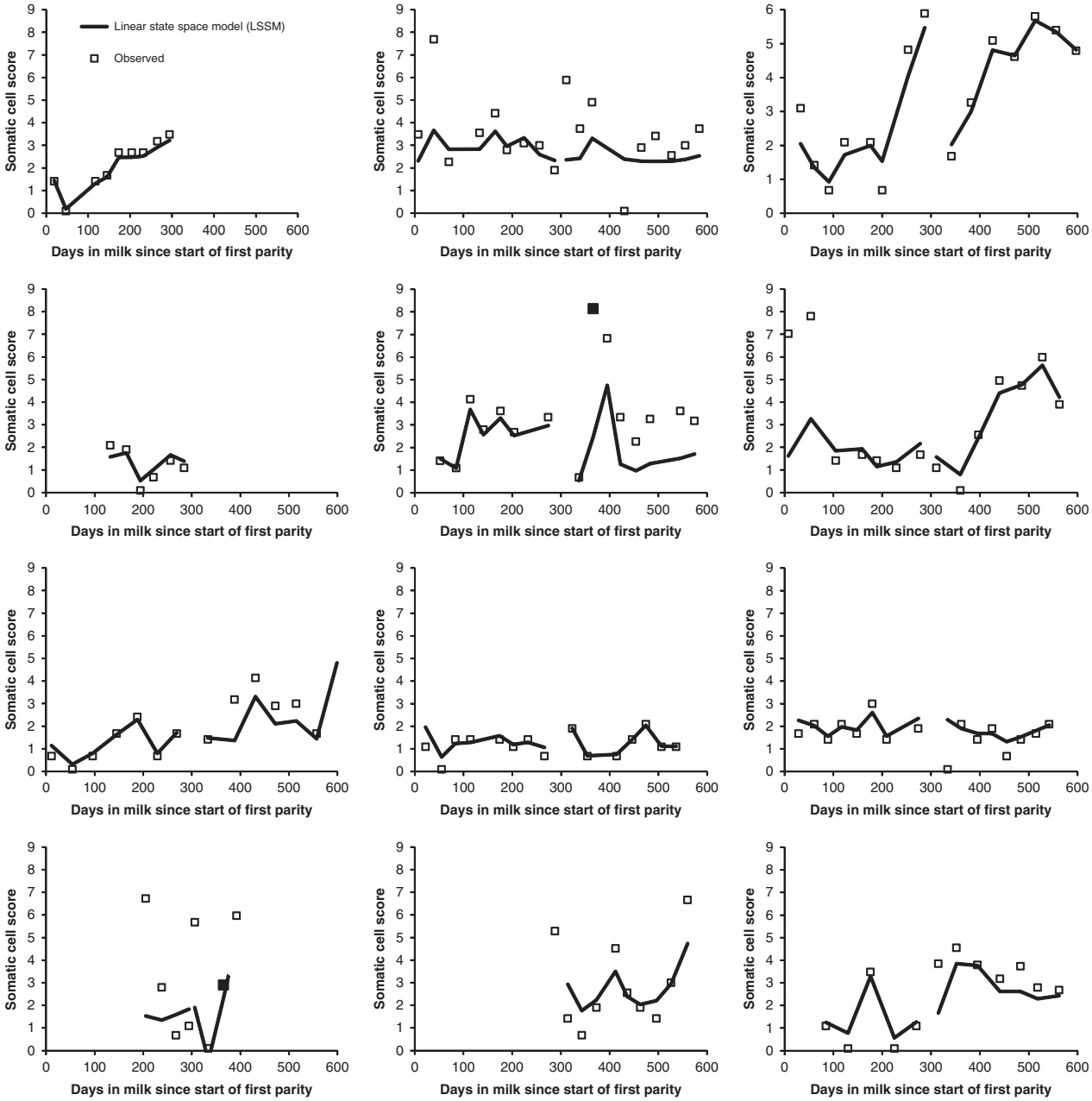

Figure 1 Observed and estimated somatic cell score per days in milk since the start of the first parity.

were reported in our study, as suggested by the observation that SCSs were higher in subclinical than clinical mastitis cases. If all clinical cases had been included in the group CM = 1 and all subclincal cases in the group CM = 2, we would have expected to have an SCS means lower in the group CM = 2 than in the group CM = 1, assuming that SCS is higher for clinical than subclinical cases.

Means and standard deviations (s.d.) of SCS (averaged over all records) were 1.56 (s.d. = 1.10), 4.47 (s.d. = 2.43) and 4.78 (s.d. = 1.26) for healthy, clinical and subclinical records, respectively. This corresponds roughly to SCC values of 60 000 in healthy, 280 000 in clinical and 340 000 cells/ml in subclinical cases. Underreporting of clinical mastitis has been evidenced, for example, in countries where clinical mastitis cases are routinely registered by farmers, such as Denmark (Klaas *et al*., 2004) and Sweden (Mork *et al*., 2009). Differences may also be because of the bacterial species responsible for clinical and subclinical cases. Indeed, de Haas *et al*. (2004) showed coliform infections have tendency to be severe with the presence of a short peak in SCC, *Staphylococcus aureus* infections are less severe (subclinical) and associated with long increased SCC, whereas *Streptococcus dysgalactiae* and *uberis* seem not strongly associated with any specific patterns of peaks in SCC.

Comparison between estimated and observed values

Figure 1 shows observed and estimated SCS lactation curves for several cows taken at random in the data set. One can see

that median values of posterior distributions are quite close to observed SCS. Correspondingly, MSE_h was very low with a mean of 1.39 over all herds.

Figure 2 is a ROC plot evaluating the discriminating ability of LSSM in estimating SCS values above thresholds from 0 to 6. The area under the ROC curve is 94.01%, which indicates that LSSM is particularly accurate in correctly classifying records.

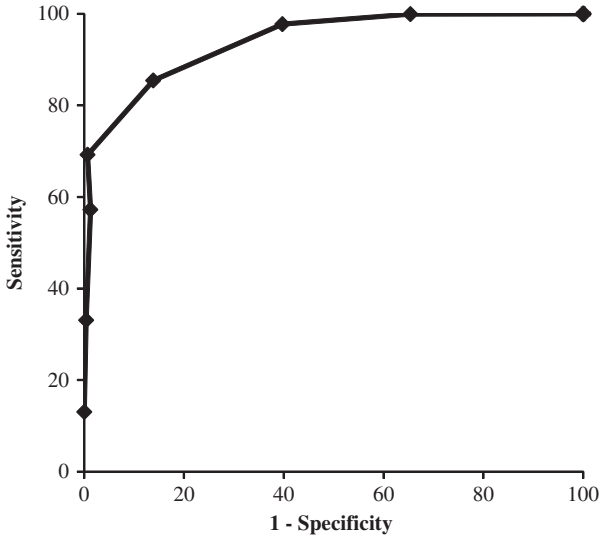

Figure 2 Receiver operating characteristic (ROC) plot.

Although it is accurate, the model is very simple and it could be made more realistic by allowing for factors known to have an effect on SCS, being at the cow or herd levels. One such set of factors would be the genetic background of the animals as numerous experiments have identified different genomic regions associated with resistance or susceptibility to mastitis (Pighetti and Elliott, 2011). Management practices (e.g. milking techniques, drying procedures, quality of feed) also influence the prevalence of specific mastitis pathogens and SCS. We could also combine SCS with measurements of biomarkers associated with the disease (Hojsgaard and Friggens, 2010).

Parameters of the model

Posterior medians and credibility intervals for all estimated parameters are shown in Figure 3 for each herd, separately. Values for h_2 are shown only for herds in which more than 2.5% of clinical cases were recorded because intervals are quite large in herds where few clinical cases were reported.

When comparing figures for the transition parameters (h_0 to h_3), one may note that h_3 estimates are slightly higher than h_1 estimates and most credibility intervals do not overlap. These estimates (divided by time interval between successive records) may be interpreted as auto-correlations between successive HID (because the transition equations correspond approximately to first-order auto-regressive processes). In our study, correlation between successive HID

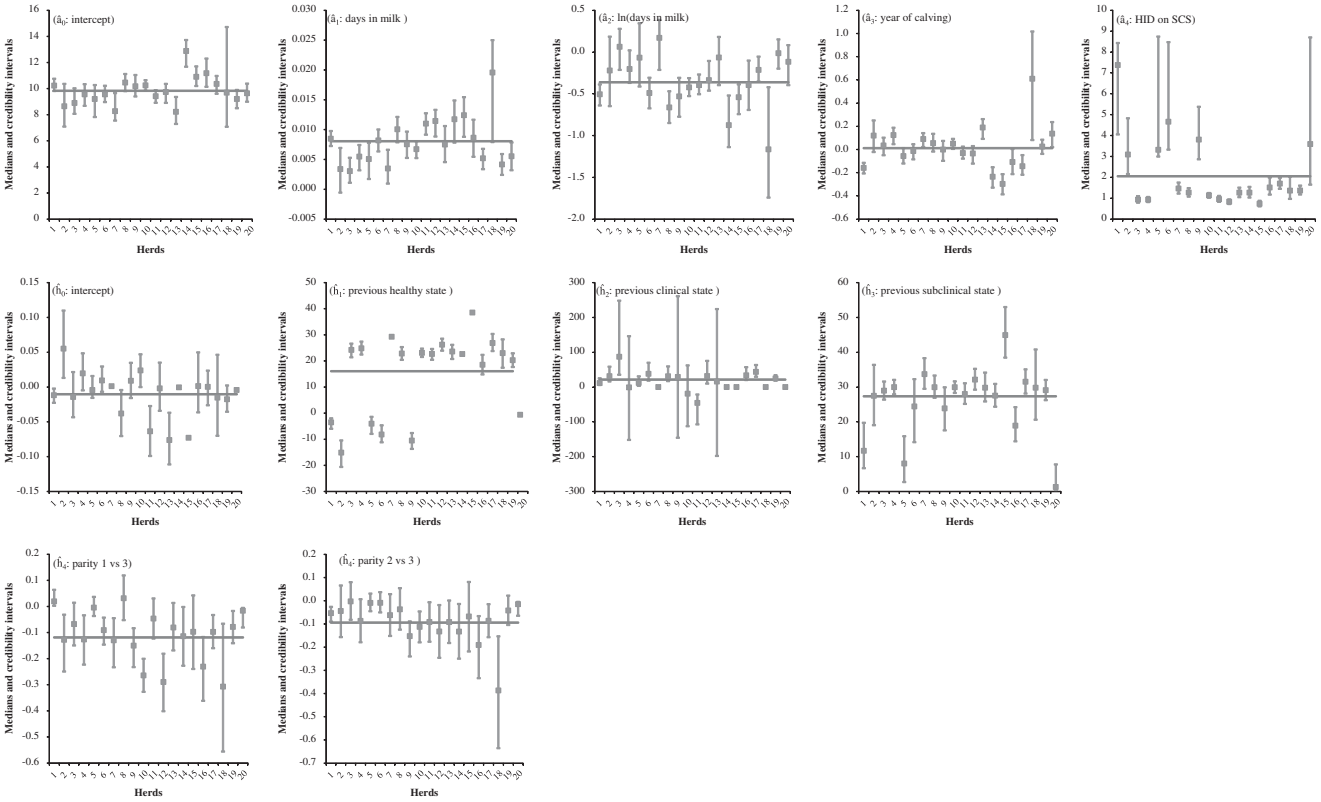

Figure 3 Posterior median and corresponding credibility intervals for the estimated model parameters.

was higher for a subclinical previous record (h_3) than a healthy previous record (h_1). Stated otherwise, this suggests HID values were linked over time when mastitis was subclinical.

The parameters may also be interpreted in light of our knowledge of the mechanisms of immune/inflammatory response to infection. Therefore, a_4 makes the link between HID and SCS and may represent our uncertainty in measuring HID (as an indicator of defence mechanisms against mastitis pathogens) with SCS. The parameters h_1 to h_3 are linked to the rates of change in leukocyte number under health and infection. In the absence of infection (and if all other factors of HID variation are kept constant), there is an equilibrium where cells enter and leave the udder. In that case, $\delta_i^t = \delta_i$ for all values of t. Then, from the transition equations, we have $\delta_i = (h_0 HSC_i^{t-1} + h_4 PAR_i^t)/(1-(h_1/LAG_i^{t-1}))$ and $h_1 = \sqrt{LAG_i^{t-1}}$ because $var(HID_i) = 1 = (h_1/LAG_i^{t-1})^2 var(HID_i)$. When pathogens are present, an extra concentration of leukocytes is recruited in the udder and cells are removed after phagocytosis and killing of pathogens. In such a situation, we may link parameters of the LSSM to parameters of ordinary differential equations (Quach *et al.*, 2007), describing the rates at which the cells enter and leave the udder. For example, if cell dynamics is modeled according to a system of equations that reproduce persistent infections (e.g. White *et al.*, 2010), then a simplified version of the system would be

$$\delta_i^t = \delta_i^0 \, e^{-h_3 t},$$

with $h_3 = (g–r)$, $g =$ rate of influx of cells into the milk in the presence of infection and $r =$ rate of removal. (We chose persistent equations so that the time frame of the LSSM is similar to the timeframe of the ordinary differential equation). This procedure, that is, the formal combination of expert knowledge in the form of dynamic modeling with LSSM on real data, is actually used for estimating disease burden based on surveillance data, often complicated by a tendency for underreporting (e.g. Hooker *et al.*, 2011).

In this paper, we propose a simple LSSM to analyze trends in SCS. The model is flexible, it corrects for measurement errors and is relatively easy to implement. It may be extended to account for the dynamics of the response to the infection. Through its hidden variable, it could help in the identification of sub-phenotypes characterized by the same level of expressed genes and decrease the gap between gene and phenotype (Wiggs, 2010). It could also lead to the identification of disease markers (Morris *et al.*, 2010) and the development of treatment to specific disease symptoms (Hall, 2013).

Acknowledgments

This study was supported by the European Research Group EADGENE_S and the Région Wallonne. The authors also thank the OSAM team that collect and store the data.

References

Ali A and Shook GE 1980. An optimum transformation for somatic cell concentration in milk. Journal of Dairy Sciences 63, 487–490.

Bradley A and Green M 2005. Use and interpretation of somatic cell count data in dairy cows. In Practice 27, 310–315.

Bradley AJ, Leach KA, Breen JE, Green LE and Green MJ 2007. Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. Veterinary Record 160, 253–257.

Chen Z and Brown EN 2013. State space models. Scholarpedia 8, 30868.

De Haas Y, Veerkamp RF, Barkema HW, Grohn YT and Schukken YH 2004. Associations between pathogen-specific cases of clinical mastitis and somatic cell count patterns. Journal of Dairy Sciences 87, 95–105.

Detilleux J and Leroy PL 2000. Application of a mixed normal mixture model for the estimation of Mastitis-related parameters. Journal of Dairy Sciences 83, 2341–2349.

Detilleux J, Theron L, Beduin JM and Hanzen C 2012. A structural equation model to evaluate direct and indirect factors associated with a latent measure of mastitis in Belgian dairy herds. Preventive Veterinary Medicine 107, 170–179.

Detilleux JC 2011. A hidden Markov model to predict early mastitis from test-day somatic cell scores. Animal 5, 175–181.

Faye B, Dorr N, Lescourret F, Barnouin J and Chassagne M 1994. Farming practices associated with the 'udder infection' complex. Veterinary Research 25, 213–218.

Gelman A and Rubin DB 1996. Markov chain Monte Carlo methods in biostatistics. Statistical Methods in Medical Research 5, 339–355.

Gianneechini R, Concha C, Rivero R, Delucci I and Moreno Lopez J 2002. Occurrence of clinical and sub-clinical mastitis in dairy herds in the West Littoral Region in Uruguay. Acta Veterinary Scandinavia 43, 221–230.

Hall IP 2013. Stratified medicine: drugs meet genetics. European Respiratory Review 22, 53–57.

Harmon RJ 1994. Physiology of mastitis and factors affecting somatic cell counts. Journal of Dairy Sciences 77, 2103–2112.

Hojsgaard S and Friggens NC 2010. Quantifying degree of mastitis from common trends in a panel of indicators for mastitis in dairy cows. Journal of Dairy Sciences 93, 582–592.

Hooker G, Ellner SP, Roditi Lde V and Earn DJ 2011. Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario. Journal of the Royal Society Interface 8, 961–974.

Hovinen M and Pyorala S 2011. Invited review: udder health of dairy cows in automatic milking. Journal of Dairy Sciences 94, 547–562.

Jamrozik J and Schaeffer LR 2010. Application of multiple-trait finite mixture model to test-day records of milk yield and somatic cell score of Canadian Holsteins. Journal of Animal Breeding and Genetics 127, 361–368.

Klaas IC, Enevoldsen C, Vaarst M and Houe H 2004. Systematic clinical examinations for identification of latent udder health types in Danish dairy herds. Journal of Dairy Sciences 87, 1217–1228.

Kristula MA, Curtis CR, Galligan DT and Bartholomew RC 1992. Use of a repeated-measures logistic regression model to predict chronic mastitis in dairy cows. Preventive Veterinary Medicine 14, 54–68.

Laevens H, Deluyker H, Schukken YH, De Meulemeester L, Vandermeersch R, De Muelenaere E and De Kruif A 1997. Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. Journal of Dairy Sciences 80, 3219–3226.

Lievaart JJ, Kremer WD and Barkema HW 2007. Short communication: comparison of bulk milk, yield-corrected, and mean somatic cell counts as parameters to summarize the subclinical mastitis situation in a dairy herd. Journal of Dairy Sciences 90, 4145–4148.

Lunn D, Spiegelhalter D, Thomas A and Best N 2009. The BUGS project: evolution, critique and future directions. Statistics in Medicine 28, 3049–3067.

Metz CE, Herman BA and Shen JH 1998. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. Statistics in Medicine 17, 1033–1053.

Miltenburg JD, de Lange D, Crauwels AP, Bongers JH, Tielen MJ, Schukken YH and Elbers AR 1996. Incidence of clinical mastitis in a random sample of dairy herds in the southern Netherlands. Veterinary Record 139, 204–207.

Mork M, Lindberg A, Alenius S, Vagsholm I and Egenvall A 2009. Comparison between dairy cow disease incidence in data registered by farmers and in data from a disease-recording system based on veterinary reporting. Preventive Veterinary Medicine 88, 298–307.

Morris AP, Lindgren CM, Zeggini E, Timpson NJ, Frayling TM, Hattersley AT and McCarthy MI 2010. A powerful approach to sub-phenotype analysis in

population-based genetic association studies. Genetic Epidemiology 34, 335–343.

Osteras O, Solbu H, Refsdal AO, Roalkvam T, Filseth O and Minsaas A 2007. Results and evaluation of thirty years of health recordings in the Norwegian dairy cattle population. Journal of Dairy Sciences 90, 4483–4497.

Pighetti GM and Elliott AA 2011. Gene polymorphisms: the keys for marker assisted selection and unraveling core regulatory pathways for mastitis resistance. Journal of Mammary Gland Biology and Neoplasia 16, 421–432.

Quach M, Brunel N and d'Alche-Buc F 2007. Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference. Bioinformatics 23, 3209–3216.

Robertson C, Sawford K, Gunawardana WS, Nelson TA, Nathoo F and Stephen C 2011. A hidden Markov model for analysis of frontline veterinary data for emerging zoonotic disease surveillance. PLoS One 6, e24833.

Reding E, Theron L, Detilleux J, Bertozzi C and Hanzen C 2012. LAECEA: un outil fédérateur d'aide à la décision pour le suivi de la santé mammaire dans les élevages bovins laitiers wallons. In the 18th annual meeting of the 3 R. Session Aide à la décision en élevage. Paris, France.

Roweis S and Ghahramani Z 1999. A unifying review of linear gaussian models. Neural Computing 11, 305–345.

Schepers AJ, Lam TJ, Schukken YH, Wilmink JB and Hanekamp WJ 1997. Estimation of variance components for somatic cell counts to determine thresholds for uninfected quarters. Journal of Dairy Sciences 80, 1833–1840.

Soyeurt H, Bastin C, Colinet FG, Arnould VM, Berry DP, Wall E, Dehareng F, Nguyen HN, Dardenne P, Schefers J, Vandenplas J, Weigel K, Coffey M, Theron L, Detilleux J, Reding E, Gengler N and McParland S 2012. Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis. Animal 6, 1830–1838.

Steeneveld W, Hogeveen H, Barkema HW, van den Broek J and Huirne RB 2008. The influence of cow factors on the incidence of clinical mastitis in dairy cows. Journal of Dairy Sciences 91, 1391–1402.

Uhler C 2009. Mastitis in dairy production: estimation of sensitivity, specificity and disease prevalence in the absence of a gold standard. Journal of Agricultural, Biological, and Environmental Statistics 14, 79–98.

White LJ, Schukken YH, Dogan B, Green L, Dopfer D, Chappell MJ and Medley GF 2010. Modelling the dynamics of intramammary E. coli infections in dairy cows: understanding mechanisms that distinguish transient from persistent infections. Veterinary Research 41, 13–21.

Wiggs JL 2010. Genotypes need phenotypes. Archives of Ophthalmology 128, 934–935.

Wolfova M, Stipkova M and Wolf J 2006. Incidence and economics of clinical mastitis in five Holstein herds in the Czech Republic. Preventive Veterinary Medicine 77, 48–64.