THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

A State Space Model Exhibiting a Cyclic Structure with an Application to Progesterone Concentration in Cow Milk

Jørgen V. Hansen, Jens L. Jensen
and Søren Højsgaard

Research Report No. 12 | September 2008

A State Space Model Exhibiting a Cyclic Structure with an Application to Progesterone Concentration in Cow Milk

Jørgen V. Hansen[*], Jens L. Jensen[†] and Søren Højsgaard[‡]

Abstract

Progesterone is a hormone linked to the reproductive status of dairy cows. Hence, with the increasing availability of on-line records of the concentration of progesterone in cow milk, there is a need for new tools to analyse such data. The aim is to find techniques for better determination of the time when cows are in oestrus to increase the rate of succesful inseminations. In this paper we propose a state space model for data with a continuous and cyclic trend in the mean. Furthermore a matching Kalman filter is developed. The model is tested on progesterone data from 112 cow-lactations with the purpose of evaluating the use of progesterone for detection of oestrus.

Keywords: cyclic model, dairy cow, Kalman filter, oestrus detection.

1 Introduction

Data from many biological processes exhibit a clear cyclic nature. A classical example is the yearly number of lynx in Canada (Elton and Nicholson, 1942) where the cyclic nature is caused by a predator-prey relationship. The example of main interest to us in this paper is one where the oestrus cycle in cows generates a cyclic behavior of the concentration of progesterone in cow milk. A short presentation of the biology behind this process is given in Section 2. In models for such data it is often natural to introduce hidden variables which have certain biological or physical interpretations. We study a state space model which generates a cyclic behavior with continuous and piecewise linear mean of the univariate observations. The continuity and cyclic nature implies that a period of increase in the mean must at some point be followed by a decrease in the mean. It can be useful to think of a model with four stages, where the four stages correspond to an increase in the mean, a high level of the mean, a decrease in the mean and finally a low level of the mean. However, this does not exhaust the variety of possible cyclic models. In this way the time axis will be divided into segments where all observations within a segment belong to

[*]University of Aarhus, Foulum and Århus, Denmark.
[†]University of Aarhus, Århus, Denmark.
[‡]University of Aarhus, Foulum, Denmark.

the same stage. The hidden variables will include one variable holding the current stage and two variables holding the time points for the beginning and end of the current stage. The inference concerning time points at which paramters change is known as changepoint detection, a subject on which Page (1954) wrote one of the first papers. The state space model we consider can be seen as a modification of the model of Fearnhead and Liu (2007) and Fearnhead and Vasileiou (2008). The precise mathematical formulation of the model is given in Section 3. In Section 4 we present an approximate filter for the hidden variables of the state space model, leaving the detailed derivation of this filter to Appendix B. In Section 5 we discuss how to estimate the parameters of the model. In Section 6 we show how a model of the class presented in Section 3 can be used to describe the level of progesterone in cows. Also in Section 6 we describe an algorithm for finding the optimal time point for insemination and we study how well the algorithm works in practice. Finally, in Section 7 we discuss the results of this analysis in the perspective of creating better tools to assess the optimal time point for artificial insemination.

2 Biological background

The main motivation for this work is to detect oestrus in cows by modelling the progesterone concentration in cow milk. Oestrus is usually defined to be the period of low progesterone in the cyclic pattern of the hormone (Peters and Ball, 1995). When we apply the cyclic model to the progesterone data we therefore are specifically interested in determining when a cow is in the stage corresponding to low progesterone measurements. Only during oestrus can the cow be succesfully inseminated and then produce a calf. Cows in oestrus usually exhibit physiological signs of sexually receptive behavior. However, the traditional, visual, detection of oestrus signs (i.e. not using progesterone) is becoming more difficult as genetic selection of cows for high milk yields has reduced the intensity of the oestrus (Dobson et al., 2008). Also, in modern dairy cow herds there is a growing need for automated management of the cows due to the large herd sizes. Therefore, detection of the time the cow is in oestrus is one such problem where a farmer could benefit from improved techniques.

The reproductive cycle of dairy cows is approximately 21 days (ranging from 18 to 26 days), and for maximal chance of success insemination should take place 12–24 hours before ovulation (Roelofs et al., 2006). Though several indicators exist for determining when cows are in oestrus (Fulkerson et al., 1983; Xu et al., 1998; Cavalieri et al., 2003) there is still a need for better prediction of the time when cows are most susceptable for insemination. Progesterone, which can now be measured automatically in the milk, is the accepted gold standard for assessing the reproductive status (Peters and Ball, 1995; Cavalieri et al., 2001; Roelofs et al., 2006).

The oestrus cycle in a cow is initiated by the creation of a follicle in the ovary. The follicle grows in size until ovulation where the follicle ruptures and releases the egg which is tranported down the oviduct toward the uterus. After the ovulation the remains of the follicle, called the corpus luteum, stay in the ovary. The cells of the corpus luteum begin to secrete progesterone approximately 4 days after ovulation. Progesterone is required for the maintance of pregnancy (Peters and Ball, 1995).

The presence of a fertilized egg (embryo) blocks regression of the corpus luteum which then continue to secrete progesterone throughout pregnancy. If the egg is not fertilized the corpus luteum will regress and stop producing progesterone approximately 17 days after ovulation. The following drop in progesterone then causes a new follicle to grow and the cycle repeats itself.

Progesterone can be measured in the milk throughout lactation which is the period in time where a cow produces milk following a calving. In this paper we define the term *cow-lactation* to be the statistical term of the cross factor of cow and parity, where parity is the number of calves a cow has given birth to. That is, a cow of first parity has just had its first calf and so on. The concentration of progesterone in the milk is measured in ng/ml and varies in the range from 0 to 30 ng/ml. An example showing the development of the concentration of progesterone in milk from a cow is seen in Figure 1. In this example the cyclic behavior of the hormone is observed through several cycles.

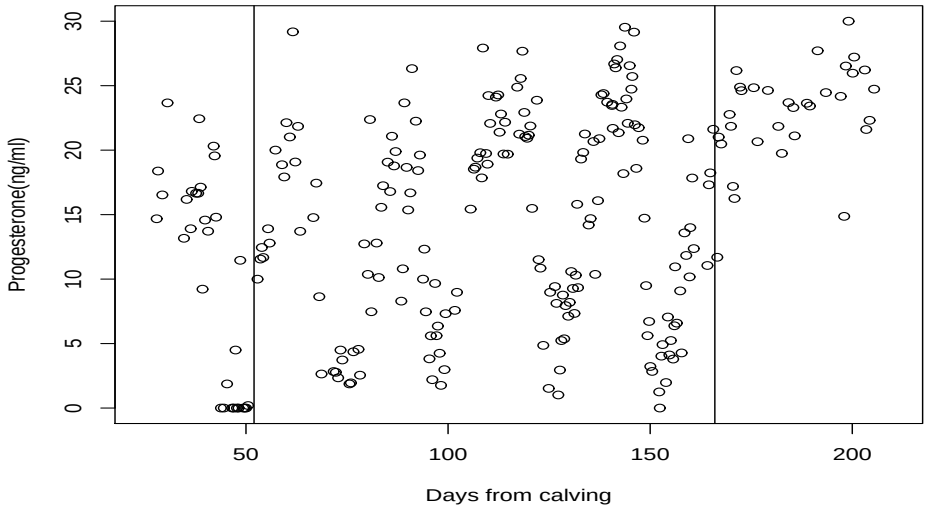

Figure 1: Progesterone measurements through a cow-lactation showing a cyclic behavior through several cycles. The model apply to data only when the cow is in its oestrus cycle. Therefore observations in the beginning and the end of lactation are excluded in the analysis. This is indicated by the vertical lines.

3 Cyclic model

In the following, the model we consider in this paper is defined. We use a state space model incorporating the idea of several different stages each describing a linear development in the mean of the observations, such that the mean as a function of time is continuous. The number of different stages is denoted by m. The m stages follow each other in the same order, so that one round of the m stages constitute a cycle of the process. More specifically, a time segment of stage $q = 1$ will be followed by a segment of stage $q = 2$, and so forth, until a segment of stage $q = m$ is again

followed by stage $q = 1$. An illustration of a possible development in the mean of the observation is shown in Figure 2. In this illustration the number of stages is $m = 4$, which is the value we use for modeling the progesterone data in Section 6.

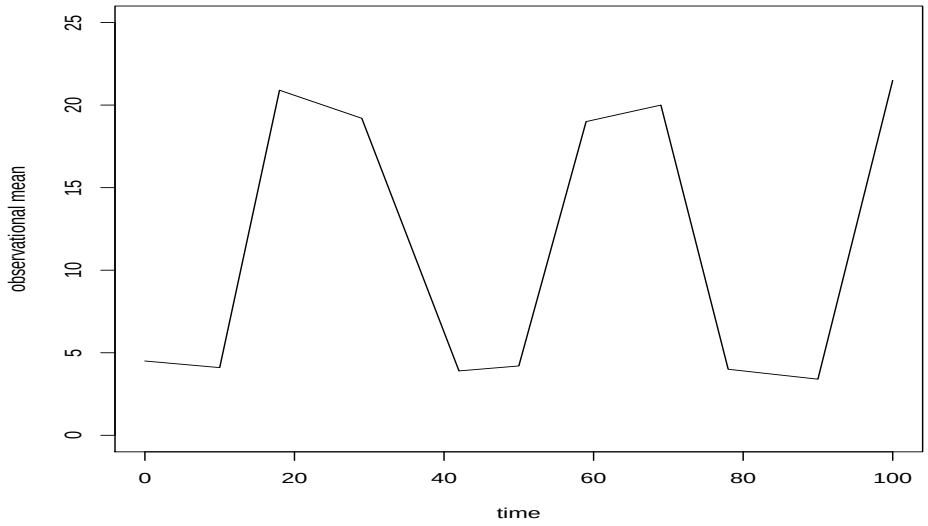

Figure 2: Example of possible development in the mean of the observations.

3.1 Hidden variables

The model has five hidden variables. The hidden process is considered in discrete time, $t \in \mathbf{Z}$. In applications these discrete times will constitute a scaling of real time. To each time point t these hidden variables contain information about the position of the change points separating the stages, and infomation on the mean level of the observations. The five hidden variables are:

R_t : the point in time prior to t with the most recent change of stage, $(R_t < t)$.

S_t : the stage entered at time R_t with value in $\{1, 2, \ldots, m\}$.

N_t : the point in time for the next change of stage after R_t, $(N_t \geq t)$.

a_t : the mean of an observation at time R_t.

b_t : the mean of an observation at time N_t.

3.2 Stochastics of the state $(R_t, S_t, N_t, a_t, b_t)$

The starting point of describing the stochastics governing the state variables is to find the positions of the changepoints. Given that there is a change to stage q at t, the distribution of the waiting time for the next change depends on q only. This

distribution is denoted W_q, where q is the new stage. The only restriction we put on W_q is that it has finite support. Formally, we write

$$(N_{t+1} - N_t | N_t = t, S_{t+1} = q) \sim W_q. \tag{3.1}$$

Throughout the paper we use the notation $W_q(r) = P(w = r)$ if $w \sim W_q$. Thus (3.1) describes the dynamics of the three discrete hidden variables (R_t, S_t, N_t). In terms of one step transition probabilities we have that

$$\begin{aligned} &P(R_{t+1} = j', S_{t+1} = q', N_{t+1} = l' | R_t = j, S_t = q, N_t = l) \\ &\quad = \begin{cases} W_{q'}(l' - t) & \text{if } l = t \text{ and } j' = t, q' = q + 1 (\text{mod } m), \\ 1 & \text{if } l > t \text{ and } j' = j < t, q' = q, l' = l, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.2}$$

The triple $(R_t - t, S_t, N_t - t)$ as defined above constitutes by itself a Markov chain. In Appendix A it is shown that the stationary distribution for $(R_t - t, S_t, N_t - t)$ is

$$\pi(j, q, l) = \frac{W_q(l - j + M)}{\sum_r \nu(r)},$$

where

$$M = \max\{r \mid \exists q : W_q(r) > 0\} \tag{3.3}$$

is the maximal possible waiting time between two consecutive changepoints and $\nu(q) = \sum_{i=1}^{\infty} i W_q(i)$ is the mean of the waiting time distribution W_q. The stationary distribution can be used as a prior when no information about the state variables is at hand at the time of the first recording.

Next, the stochastic behavior of the continuous hidden variables a_t and b_t is described. These two variables hold information about the mean of the observations within a stage. For each changepoint t, let $x(t)$ denote the hidden mean of a possible observation y_t. If the stage of a time interval beginning at time t is q, then we assume

$$(x(t) | N_t = t, S_{t+1} = q) \sim N(\mu_q, \omega_q^2), \tag{3.4}$$

where μ_q and ω_q^2 are parameters. The hidden variables a_t and b_t are then defined to be

$$a_t = x(R_t) \text{ and } b_t = x(N_t). \tag{3.5}$$

That is, a_t is the mean at the beginning and b_t the mean at the end of the stage entered at time R_t. This means that if there is no change point at time t then $(a_{t+1}, b_{t+1}) = (a_t, b_t)$. On the other hand if there is a change of stage at time t, then $a_{t+1} = b_t$ and $b_{t+1} = x(N_{t+1})$.

3.3 Stochastics of observations

The mean at any time point $s \in \mathbf{R}$ we define by linear interpolation using the mean at the end points of the stage. That is,

$$x(s) = a_t + \frac{b_t - a_t}{N_t - R_t}(s - R_t), \quad s \in \mathbf{R}, t = \lceil s \rceil,$$

where $\lceil s \rceil$ is the smallest integer greater than or equal to s. In this way the underlying mean $x(s)$ of the observation given the hidden variables is continuous and piecewise linear.

Data $\{(y_i, s_i) | i = 1, \ldots, n\}$ consist of a set of observations y_i recoded at time points s_i, where n is the number of observations. To define the distribution of data, assume that we have an observation y_i at time $s_i \in \mathbf{R}$. Note that we do not restrict the observations to occur at time points that are multiples of the time unit. If the stage at time s_i is q, that is, $S_{\lceil s_i \rceil} = q$, we assume

$$\left(y_i | (R, S, N, a, b)_{\lceil s_i \rceil}\right) \sim N(x(s_i), \sigma_q^2),$$

where $\sigma_q^2, q = 1, 2, \ldots, m$ are parameters.

To summarize the parameters of the model are

μ_q : the mean of the hidden stochastic mean at a time point where the stage changes to q,

ω_q^2 : the variance of the hidden stochastic mean at a time point where the stage changes to q,

σ_q^2 : residual variance of observations within stage q,

for $q = 1, 2, \ldots, m$. Please note that the waiting time distributions $W_q, q = 1, 2, \ldots, m$ may depend on an unknown parameter θ.

4 An approximate filter

An approximate filter for the state space model presented in Section 3 is described below. By using the notation $y^s = \{y_i \mid s_i \leq s\}$ and $y_r^s = \{y_i \mid r < s_i \leq s\}$ our goal is to determine the filter densities which we write as

$$p(R_t = j, S_t = q, N_t = l, a_t, b_t | y^t) = p_t(j, q, l, a_t, b_t) \text{ for all t } \in \mathbf{N}. \tag{4.1}$$

We use the approximation

$$p_t(j, q, l, a_t, b_t) = p_t(j, q, l) \phi(a_t, b_t;\ \mu_t(j, q, l), \Sigma_t(j, q, l)), \tag{4.2}$$

where $p_t(\cdot, \cdot, \cdot)$ on the right hand side of (4.2) is the marginal density of (R_t, S_t, N_t) and $\phi(\cdot, \cdot;\ \mu, \Sigma)$ is the normal density with mean μ and variance Σ. Therefore the filter densities (4.1) are specified by $p_t(j, q, l)$, $\mu_t(j, q, l)$ and $\Sigma_t(j, q, l)$. Here we only state the recursions for updating the filter. A detailed derivation of the filter is given in Appendix B.

Let $\sum_{(t)}$ (and $\prod_{(t)}$) denote the sum (and the product) over the set of observations $\{y_i\}$ in the time interval from t to $t+1$. When $j = R_{t+1} < t$ the filter recursions are given as follows:

$$\begin{aligned} &p_{t+1}(j, q, l) \\ &\quad = c_{t+1}(y^{t+1}) p_t(j, q, l) \frac{\phi(0;\ \mu_t(j, q, l), \Sigma_t(j, q, l))}{\phi(0;\ \mu_{t+1}(j, q, l), \Sigma_{t+1}(j, q, l))} \prod_{(t)} \phi(y_i;\ 0, \sigma_q^2), \end{aligned} \tag{4.3}$$

with

$$\Sigma_{t+1}(j,q,l)^{-1} = \Sigma_t(j,q,l)^{-1} + \frac{1}{\sigma_q^2(l-j)^2}\sum_{(t)}\begin{pmatrix} (l-s_i)^2 & (s_i-j)(l-s_i) \\ (s_i-j)(l-s_i) & (s_i-j)^2 \end{pmatrix} \quad (4.4)$$

and

$$\Sigma_{t+1}(j,q,l)^{-1}\mu_{t+1}(j,q,l) = \Sigma_t(j,q,l)^{-1}\mu_t(j,q,l) + \frac{1}{\sigma_q^2}\sum_{(t)}\begin{pmatrix} y_i(l-s_i)/(l-j) \\ y_i(s_i-j)/(l-j) \end{pmatrix}. \quad (4.5)$$

The constant of proportionality $c_{t+1}(y^{t+1})$ is found from (4.3) and (4.6) below together with the condition $\sum_{j,q,l} p_{t+1}(j,q,l) = 1$.

When $R_{t+1} = t$ the recursions are

$$p_{t+1}(t,q,l) = c_{t+1}(y^{t+1})W_q(l-t)\prod_{(t)}\phi(y_i;\ 0,\sigma_q^2)\sum_{j'<t}p_t(j',q',t) \times \frac{\phi(0;\ \mu_t(j',q',t)_2, \Sigma_t(j',q',t)_{22})\phi(0;\ \mu_{\tilde{q}},\omega_{\tilde{q}}^2)}{\phi(0;\ \bar{\mu}_t(j',q',t,l), \bar{\Sigma}_t(j',q',t,l))}, \quad (4.6)$$

$$\mu_{t+1}(t,q,l) = \sum_{j'<t}\alpha_t(j',q',t,l)\bar{\mu}_t(j',q',t,l), \quad (4.7)$$

and

$$\Sigma_{t+1}(t,q,l) = \sum_{j'<t}\alpha_t(j',q',t,l)[\bar{\Sigma}_t(j',q',t,l) + \bar{\mu}_t(j',q',t,l)\bar{\mu}_t(j',q',t,l)^T] - \mu_{t+1}(t,q,l)\mu_{t+1}(t,q,l)^T \quad (4.8)$$

where $\tilde{q} = q+1 (\text{mod } m)$ and where $\bar{\mu}_t(j',q',t,l)$, $\bar{\Sigma}_t(j',q',t,l)$ and $\alpha_t(j',q',t,l)$ are given in Appendix B.

5 Parameter estimation

Maximum likelihood can be used to estimate the residual variances $\sigma_q^2, q = 1,\ldots,m$. This estimation procedure is described in Section 5.1. The waiting time distributions can be estimated using an EEE-algorithm as described in Section 5.2. We do not suggest any general procedures for estimation of the parameters μ_q and ω_q^2, $q = 1,\ldots,m$, but in Section 6 we describe how to find crude estimates of these parameters to use for the modelling of the progesterone data.

5.1 Estimation of the residual variance

The residual variance parameters σ_q^2, $q = 1,\ldots,m$ are estimated using maximum likelihood. The constant of proportionality $c_{t+1}(y^{t+1})$ in (4.3) and (4.6) can, from the derivation of the filter in Appendix B, be seen to be

$$c_{t+1}(y^{t+1}) = \frac{p(y^t)}{p(y^{t+1})} = \frac{1}{p(y_t^{t+1}|y^t)}.$$

Therefore, using the approximate filter, we can calculate an approximation to the likelihood function

$$L(\sigma_1^2, \ldots, \sigma_m^2) = p(y^n) = \prod_t p(y_t^{t+1}|y^t),$$

which can be maximized using numerical techniques to find estimates $\hat{\sigma}_q^2$ of the residual variances.

5.2 Estimation of the distribution of waiting times

Given M as defined in (3.3) and a model for the waiting time distributions W_q, $q = 1, \ldots, m$, we can estimate the parameter θ of this model using an EEE algorithm which is proposed and discussed in e.g. Heyde and Morton (1996), Rosen et al. (2000) and Elashoff and Ryan (2004). Fundamentally, an EEE algorithm works similar to an EM algorithm (Dempster et al., 1977). The difference is that the M-step of maximizing the likelihood is replaced by a step where an estimating equation is solved. In the special case where the estimating equation is the likelihood equation the EEE algorithm is an EM algorithm. For a parameter θ of the waiting time distribution we use an estimating function of the form $\sum_1^n \psi_i$, where $\psi_i = \psi(z_i, z_{i-1}; \theta)$ for a function ψ, and where $z_i = (R_i, S_i, N_i)$. The E (expectation) step is to calculate

$$E\Big(\sum_1^n \psi_i | y^n\Big),$$

where n is the number of observations. In Appendix C it is shown that $E\big(\sum_1^k \psi_i | y^k\big)$ can be calculated iteratively in k using the filter probablities of Section 4. As an example consider the case where the waiting time probabilities are modelled with no other restriction than $\sum_{l=1}^M W_q(l) = 1$ for all q. We can then use the estimating functions

$$\begin{aligned}&\psi(z_t, z_{t-1}; q, l)\\ &\quad = 1(R_t = t-1, S_t = q, N_t = t-1+l) - W_q(l)1(R_t = t-1, S_t = q),\end{aligned}$$

where $1(\cdot)$ is the indicator function. In the EE (estimating equation) step the new value of $W_q(l)$ becomes

$$W_q(l) = \frac{E\left(\sum_1^n 1(R_t = t-1, S_t = q, N_t = t-1+l)|y_1^n\right)}{E\left(\sum_1^n 1(R_t = t-1, S_t = q)|y_1^n\right)},$$

where the nominator and denominator have been found in the E step.

6 Application

6.1 Progesterone data

The objective of this part of the study is to test the ability of the model to predict the time of oestrus in cows from on-line progesterone concentration measurements. We consider a data set where progesterone measurements were made on milk

samples taken from all milking cows in one research herd (Danish Cattle Research Centre) during the period 12 Sept. 2002 to 30 Sept. 2006. In the this dataset 123 cow-lactations included an oestrus that was identified as a confirmed oestrus, i.e. an oestrus at which insemination resulted in a confirmed pregnancy. A detailed description of the collection of data can be found in Friggens et al. (2008).

Parts of the dataset were collected before the cows had entered their oestrus cycle or after the cows had succesfully been inseminated as seen in Figure 1. Because only data collected when the cows are in their oestrus cycle are of interest, we excluded parts of the data at the beginning and at the end of each cow-lactation. For a small number of cows no data was left by this reduction of data. This was primarily the case when the cow was succesfully inseminated at the first oestrus which is not preceeded by a high progesterone stage. This reduction left us with 112 cow-lactations.

6.2 Model for progesterone data

The concentration of progesterone in milk has a cyclic nature with an average cycle length of about 21 days (ranging from 18 to 26 days). There is from cycle to cycle a small variation in the cycle length within cows. Roughly the cyclic nature of the progesterone content can be described in the following way. In each cycle we see four different stages for the concentration of progesterone. Each with a different time length. The four stages we enumerate as follows: 1. Low level of progesterone, 2. Slow increase in progesterone, 3. High level of progesterone, 4. Rapid decrease in progesterone.

It seems reasonable to assume that the mean level at the beginning and at the end of a low stage or a high stage is roughly the same. According to the model specification in (3.4) we can formulate this as the restriction $\mu_1 = \mu_2$ and $\mu_3 = \mu_4$. Also we let $\omega_1^2 = \omega_2^2$ and $\omega_3^2 = \omega_4^2$. The parameters μ_q and ω_q^2, $q = 1, 2, 3, 4$ will be the same for all cows and we estimate them as described in Section 6.3.1.

We model the waiting time distributions $W_q, q = 1, 2, 3, 4$ as discretized gamma distributions truncated at M. The gamma distributions are parameterized with parameters α_q and β_q, such that the means and variances are α_q/β_q and α_q/β_q^2, respectively. The waiting time distributions are also asumed to be the same for all cows. To estimate the α_q's and β_q's we use the estimating functions

$$\begin{aligned}\psi_1(z_t, z_{t-1}) = 1(R_t = t-1, S_t = q)\log\beta_q - 1(R_t = t-1, S_t = q)\frac{\Gamma'(\alpha_q)}{\Gamma(\alpha_q)} \\ + \sum_l \log(l)\cdot 1(R_t = t-1, S_t = q, N_t = t-1+l) \qquad (6.1)\end{aligned}$$

and

$$\begin{aligned}\psi_2(z_t, z_{t-1}) = 1(R_t = t-1, S_t = q)\frac{\alpha_q}{\beta_q} \\ - \sum_l l\cdot 1(R_t = t-1, S_t = q, N_t = t-1+l) \qquad (6.2)\end{aligned}$$

chosen so as to resemble the likelihood equations for a sample from a gamma distribution.

Finally we take the residual variances to be the same for all four stages, but specific for each cow. The cow specific variances are estimated as described in Section 5.1.

6.3 Result of analysis

6.3.1 Estimation of parameters

We estimate the parameters μ_q and σ_q^2, $q = 1, \ldots, 4$ in the following way. For each cow-lactation we find the 85 percent quantile of all observations in that cow-lactation and regard this as an outcome of a $N(\mu_3, \omega_3^2)$-distributed variable. Similarly, we regard the 15 percent quantile as an outcome of a $N(\mu_1, \omega_1^2)$-distributed variable. In Figure 3, the 15 percent and the 85 percent quantile of all progesterone measurements for the cow-lactation of Figure 1 is indicated by horizontal lines. Calculating the mean and variance of these quantiles from all cow-lactations we obtain estimates for the μ_q's and the ω_q^2's. These estimates will then be used as inputs when we make inference using the filter. The estimates are given in the second and third column of Table 1.

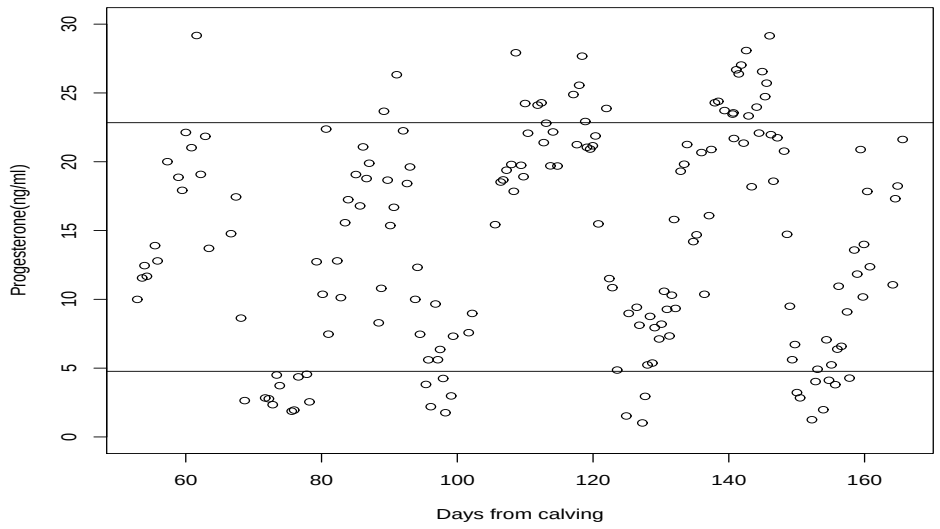

Figure 3: In this plot the 15 percent and the 85 percent quantile of all measurements are indicated by the horizontal lines. These two quantiles are used to create crude estimates of the parameters μ_q and ω_q^2, $q = 1, 2, 3, 4$.

The procedure for estimating the residual variances are carried out for fixed values of the parameters α_q and β_q of the waiting time distributions and vice versa. Therefore an iterative procedure is used for the simultaneous estimation where each of the two sets of parameters $(\alpha_q, \beta_q, q = 1, 2, 3, 4)$ and (σ^2 specific for every cow-lactation) are updated one at a time. To speed up the computations we take the unit of time to be one day during the estimation of the parameters. Furthermore, the maximal length of the waiting time is set to $M = 12$ days. The estimates for the four

Table 1: Estimates of those parameters of the cyclic model that are the same for all cow-lactations.

q	μ_q (ng/ml)	ω_q (ng/ml)	$E(W_q)$ (days)	$\sqrt{V(W_q)}$ (days)
1	3.122	1.170	5.12	2.27
2			8.25	1.81
3	20.929	2.074	6.82	2.80
4			2.74	0.98

gamma distributions defining the waiting times are given in Table 1 summarized as means (α/β) and standard deviations ($\sqrt{\alpha/\beta^2}$).

The result of estimating the residual variances is summarized in a histogram of the standard deviations shown in Figure 4.

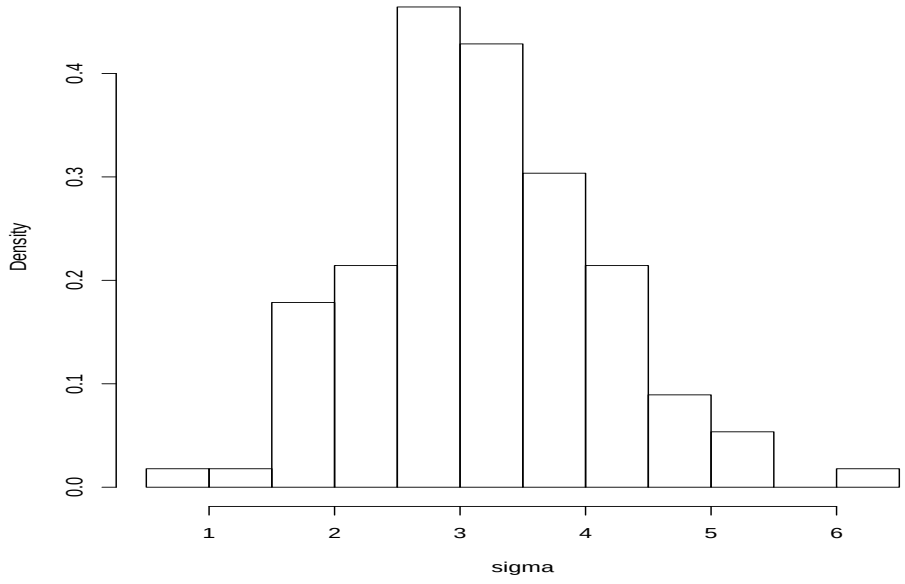

Figure 4: Histogram of the estimates of the cow-lactation specific standard deviation σ.

6.3.2 Prediction of oestrus

For each cow-lactation in our data the day (but not the exact hour) at which an artificial insemination resulted in a confirmed pregnancy is known. The filter was run with a 6-hour interval between the updates. In Figure 5 the development of the filter probability of being in the low stage, P(S=1), is shown for nine cow-lactations. In each plot a vertical line is drawn at noon on the day of succesful artificial insemination.

Knowing the day for the confirmed succesful inseminations we have the possibility

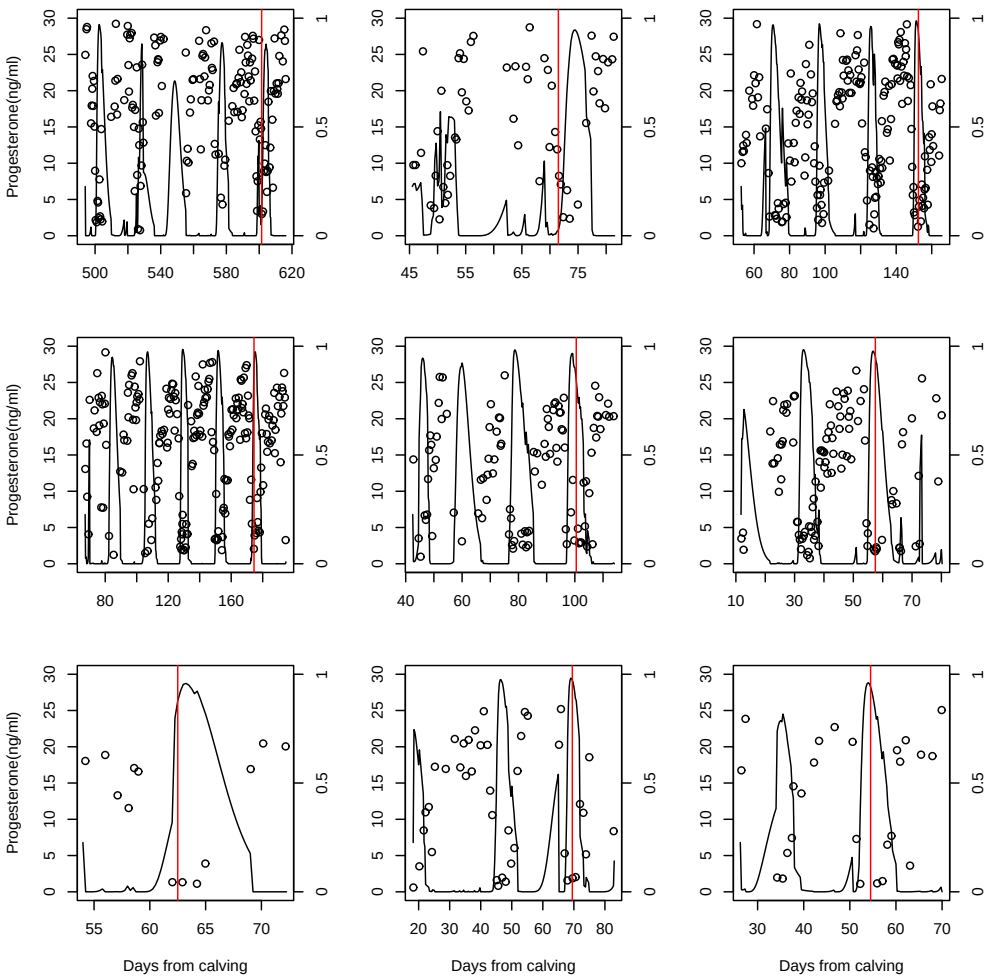

Figure 5: Filter probabilities (full drawn curve) of being in the low stage plotted against time for nine cow-lactations. The probability scale is shown on the right vertical axis. The vertical line shows the time of an insemination resulting in a confirmed pregnancy.

of evaluating how well our model can predict the time point when a cow enters oestrus. A possible way of constructing an alarm telling the farmer that a cow is about to go into oestrus is to say that when the probability of being in the low stage increases to a certain level the alarm should go off. For most of the cow-lactations the alarm goes off more than once because we observe more than one cycle for most cows. In this case we take the time of alarm t_{predict} to be the time point closest to the stipulated time of confirmed succesful insemination t_{ins}, which we in all cases define to be at noon. The difference $t_{\text{ins}} - t_{\text{predict}}$ indicates how much time in advance the farmer is given to observe the cow in detail. For 7 out of the 112 cow-lactations, the time point for insemination was placed outside the time range of the observations (which in Figure 1 is outside the two vertical lines) and therefore no t_{predict}-value was found. With a threshold probability of 0.5 of $P(S = 1)$ we observed that the this probability did not exceed the threshold in the cycle including the time of insemination for 3 of the remaining 105 cow-lactations. This leave 102 values of $t_{\text{ins}} - t_{\text{predict}}$. In Figure 6 a histogram of these values are shown. The observed mean

and standard deviation of this sample was 1.431 days and 1.556 respectively meaning that on average the alarm will tell the farmer to look for signs of oestrus a little less than one and a half day before oestrus actually occurs. In 88 of the 102 cases the alarm went off prior to the actual insemination of these cows.

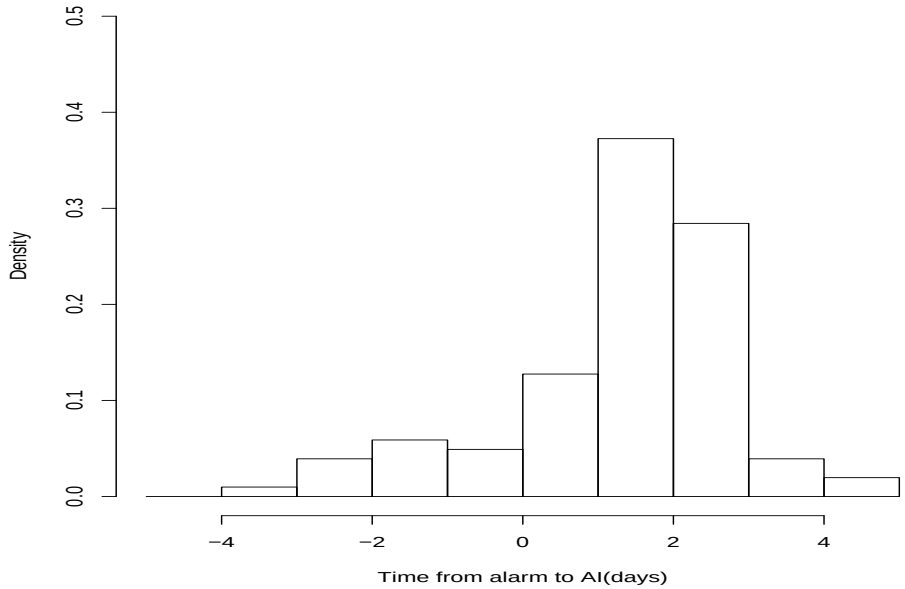

Figure 6: Histogram of the time from alarm goes off to known time for artificial insemination $t_{\text{ins}} - t_{\text{predict}}$. The observed mean of this sample is 1.431 meaning that on average the alarm will tell the farmer to look for signs of oestrus a little less than one and a half day before oestrus actually occurs.

All analysis was performed using R (R Development Core Team, 2008). The speed of running the whole filter is proportional to the cube of the number of updates per day. If for a cow-lactation the filter is updated each hour with $M = 288$ hours (12 days) an update takes approximately 200 seconds on a standard laptop.

7 Conclusions

The first objective of this study was to develop a state space model with a corresponding Kalman filter to model data with a cyclic nature. This has been done as described in Section 3 and Section 4. Furthermore in Section 5 we discussed techniques for estimation of some of the parameters in the model.

In Section 6 we analysed the progesterone data using the state space model developed in this study. We discussed how the model was able to provide an alarm for oestrus in cows. Because the aim was to evalute the use of progesterone for detection of oestrus, the time difference $t_{\text{ins}} - t_{\text{predict}}$ has to be positive so that the farmer is told to look for signs of oestrus before oestrus occurs. To be an efficient mechanism for detection of oestrus the variance of $t_{\text{ins}} - t_{\text{predict}}$ must be as small as possible. In Section 6 we found the mean of $t_{\text{ins}} - t_{\text{predict}}$ to be positive fulfilling the first requirement of a possible alarm. To judge if the corresponding variance is small enough, for this study to prove the usefulness of progesterone in oestrus

detection, two issues with the data need to be mentioned. Firstly, only the day for succesful insemination is given for each cow-lactation. Secondly, the time point of the succesful insemination is not the optimal measure to evaluate an alarm. We would rather wish to know the time at which an insemination has the highest probability of being succesful. In biological terms, this is related to the time of ovulation, which has been shown to be a rather variable time interval after the onset of oestrus. Onset of oestrus can not be measured by progesterone. The confirmed inseminations in our data are spread around this time of highest probability of success with some variation. Both of these issues contribute to a certain variance that no alarm based on any progesterone model can remove.

8 Acknowledgements

We gratefully acknowledge the contribution of the farm staff at the Danish Cattle Research Centre for providing the data used in this study, which was partly funded by the Danish Ministry of Food, Agriculture and Fisheries and the Danish Cattle Association through the Biosens project. Further, the first author thanks Paul Fearnhead for inspiration in the process of constructing the cyclic model and the corresponding filter.

A Stationary distribution of (R_t, S_t, N_t)

Lemma 1. *The stationary distribution of* $(R_t - t, S_t, N_t - t)$ *is*

$$\pi(j, q, l) = \frac{W_q(l - j)}{\sum_r \nu(r)}.$$

Proof. With

$$P(j, q, l|j', q', l') = P(R_{t+1} = j, S_{t+1} = q, N_{t+1} = l|R_t = j', S_t = q', N_t = l')$$

we must show that

$$\sum_{j',q',l'} \pi(j', q', l')P(j, q, l|j', q', l') = \pi(j, q, l),$$

for all j, q and l. We split the proof in two cases. Firstly we consider the case of no changepoint at time t which means $j < -1$. Then by (3.2) $P(j, q, l|j', q', l') = 0$ unless $j' = j + 1$, $l' = l + 1$ and $q' = q$. Also by (3.2) we find that

$$\pi(j + 1, q, l + 1)P(j, q, l|j + 1, q, l + 1) = \pi(j, q, l).$$

In the other case where $j = -1$ corresponding to a changepoint at time t we have that $P(j, q, l|j', q', l') = 0$ unless $q' = q - 1(\text{mod } m)$ and $l' = 0$. Here we find that

$$\begin{aligned}
&\sum_{j'} \pi(j', q - 1(\text{mod } m), 0)P(-1, q, l|j', q - 1(\text{mod } m), 0) \\
&\quad = \sum_{j'} \frac{W_{q-1(\text{mod } m)}(0 - j')}{\sum_r \nu(r)} W_q(l + 1) \\
&\quad = \frac{1}{\sum_r \nu(r)} W_q(l - j),
\end{aligned}$$

to complete the proof of the lemma. □

B Mathematical description of updating equations

In section 4 we presented an approximate filter for the state variables of our model. Here we give a detailed derivation of the filter recursions. That is, we derive $p(R_{t+1} = j, S_t = q, N_{t+1} = l, a_{t+1}, b_{t+1}|y^{t+1}) = p_{t+1}(j, q, l, a_{t+1}, b_{t+1})$ from p_t, the transition density and the likelihood of y_t^{t+1}, using standard updating formulas. Because we use the approximation (4.2), at each time point t we have to update the quantities $p_t(j, q, l)$, $\mu_t(j, q, l)$ and $\Sigma_t(j, q, l)$ for all j, q and l. Therefore we now assume that all the quantities are known at time t and in the following we will prove that the updating equations given in Section 4 are valid.

We first consider the case where $j < t$, that is, there is no change of the stage at time t. In this case according to (3.2), $R_t = R_{t+1}$, $S_t = S_{t+1}$ and $N_t = N_{t+1}$. Then

by (3.5), $a_t = a_{t+1}$ and $b_t = b_{t+1}$. Therefore only one transition is possible and we get directly

$$\begin{aligned}
&p_{t+1}(j,q,l,a,b)\\
&\quad = c_{t+1}(y^{t+1})p_t(j,q,l)\phi(a,b;\ \mu_t(j,q,l),\Sigma_t(j,q,l))\\
&\qquad \times \prod_{(t)} \phi\Big(y_i;\ a\frac{l-s_i}{l-j} + b\frac{s_i-j}{l-j}, \sigma_q^2\Big)\\
&\quad = p_t(j,q,l)\frac{\phi(0;\ \mu_t(j,q,l),\Sigma_t(j,q,l))}{\phi(0;\ \tilde{\mu}_t(j,q,l),\tilde{\Sigma}_t(j,q,l))}\\
&\qquad \times \phi(a,b;\ \tilde{\mu}_t(j,q,l),\tilde{\Sigma}_t(j,q,l))\prod_{(t)}\phi(y_i;\ 0,\sigma_q^2), \qquad \text{(B.1)}
\end{aligned}$$

with $\tilde{\Sigma}_t(j,q,l)^{-1}$ and $\tilde{\mu}_t(j,q,l)$ given by the right hand sides of (4.4) and (4.5). The normalizing constant $c_{t+1}(y^{t+1})$ is $p(y^t)/p(y^{t+1})$. Integrating (B.1) with respect to (a,b) the $\phi(a,b;\cdot)$ term disapear, and the formula (4.3) for $p_{t+1}(j,q,l)$ is obtained. Next dividing (B.1) by (4.3) we see that the filtering distribution of (a_{t+1},b_{t+1}) is the normal distribution with mean $\tilde{\mu}_t(j,q,l)$ and variance $\tilde{\Sigma}_t(j,q,l)$ which proves (4.4) and (4.5).

We next consider the case where $j = t$, that is, there is a change of the stage at time t. In this case $N_t = t$, R_t can be any value $j' < t$, $S_t = q' = q-1(\text{mod } m)$, and $b_t = a_{t+1} = a$. Letting $\tilde{q} = q+1(\text{mod } m)$, using the transistion density (3.2) we find

$$\begin{aligned}
p_{t+1}(t,q,l,a,b) &= c_{t+1}(y^{t+1})W_q(l-t)\sum_{j'<t} p_t(j',q',t)\\
&\quad \times \int_{a'} \phi(a',a;\ \mu_t(j',q',t),\Sigma_t(j',q',t))\phi(b;\ \mu_{\tilde{q}},\omega_{\tilde{q}}^2)\\
&\quad \times \prod_{(t)} \phi(y_i;\ a\frac{l-s_i}{l-t} + b\frac{s_i-t}{l-t},\sigma_q^2)\\
&= W_q(l-t)\sum_{j'<t} p_t(j',q',t)\phi(a;\ \mu_t(j',q',t)_2,\Sigma_t(j',q',t)_{22})\\
&\quad \times \phi(b;\ \mu_{\tilde{q}},\omega_{\tilde{q}}^2)\prod_{(t)}\phi(y_i;\ a\frac{l-s_i}{l-t} + b\frac{s_i-t}{l-t},\sigma_q^2)\\
&= W_q(l-t)\prod_{(t)}\phi(y_i;\ 0,\sigma_q^2)\sum_{j'<t} p_t(j',q',t)\\
&\quad \times \frac{\phi(0;\ \mu_t(j',q',t)_2,\Sigma_t(j',q',t)_{22})\phi(0;\ \mu_{\tilde{q}},\omega_{\tilde{q}}^2)}{\phi(0;\ \bar{\mu}_t(j',q',t,l),\bar{\Sigma}_t(j',q',t,l))}\\
&\quad \times \phi(a,b;\ \bar{\mu}_t(j',q',t,l),\bar{\Sigma}_t(j',q',t,l)), \qquad \text{(B.2)}
\end{aligned}$$

with

$$\begin{aligned}
\bar{\Sigma}_t(j',q',t,l)^{-1} &= \begin{pmatrix} \Sigma_t(j',q',t)_{22}^{-1} & 0\\ 0 & (\omega_{\tilde{q}}^2)^{-1}\end{pmatrix}\\
&\quad + \frac{1}{\sigma_q^2}\sum_{(t)}\begin{pmatrix} [(l-s_i)/(l-t)]^2 & (s_i-t)(l-s_i)/(l-t)^2\\ (s_i-t)(l-s_i)/(l-t)^2 & [(s_i-t)/(l-t)]^2\end{pmatrix} \qquad \text{(B.3)}
\end{aligned}$$

and

$$\bar{\Sigma}_t(j',q',t,l)^{-1}\bar{\mu}_t(j',q',t,l)$$
$$= \begin{pmatrix} \mu_t(j',q',t)_2/\Sigma_t(j',q',t)_{22} \\ \mu_{\tilde{q}}/\omega_{\tilde{q}}^2 \end{pmatrix} + \frac{1}{\sigma_q^2}\sum_{(t)} \begin{pmatrix} y_i(l-s_i)/(l-t) \\ y_i(s_i-t)/(l-t) \end{pmatrix} \qquad \text{(B.4)}$$

Integrating (B.2) with respect to (a,b) the $\phi(a,b;\cdot)$ term disappear, and the formula (4.6) is obtained. Dividing (B.2) by (4.6) we see that the density of (a_{t+1},b_{t+1}) is

$$\sum_{j'<t}\alpha_t(j',q',t,l)\phi(a,b;\bar{\mu}_t(j',q',t,l),\bar{\Sigma}_t(j',q',t,l)), \qquad \text{(B.5)}$$

where

$$\alpha_t(j',q',t,l) = \frac{\gamma_t(j',q',t,l)}{\sum_{\tilde{j}<t}\gamma_t(\tilde{j},q',t,l)}$$

with

$$\gamma_t(j',q',t,l) = p_t(j',q',t)\frac{\phi(0;\ \mu_t(j',q',t)_2,\Sigma_t(j',q',t)_{22})}{\phi(0;\ \bar{\mu}_t(j',q',t,l),\bar{\Sigma}_t(j',q',t,l))}.$$

We approximate the Gaussian mixture in (B.5) by a single Gaussian density with the same mean and variance. This gives the formulas (4.7) and (4.8).

C Estimation of waiting time distributions

When we estimate the waiting time probabilities $W_q(l)$ as described in Section 5.2 we need to calculate $E(\sum_1^k \psi_i|y^k)$ iteratively. Now let $\psi_i = \psi(z_i,z_{i-1})$ be a general estimation function where $z_i = (R_i,S_i,N_i)$. Using the approximation

$$p(z_{k+1},y_{k+1}|z_1^k,y_1^k) \approx p(z_{k+1},y_{k+1}|z_k,y_1^k)$$

we have the following updating rule

$$E\Big(\sum_1^{k+1}\psi_i|z_{k+1},y_1^{k+1}\Big)$$
$$= \sum_{z_k}\Big(E(\sum_1^k\psi_i|z_k,y_1^k)+\psi_{k+1}\Big)\frac{p(z_{k+1},y_{k+1}|z_k,y_1^k)p(z_k|y_1^k)}{\sum_{\tilde{z}_k}p(z_{k+1},y_{k+1}|\tilde{z}_k,y_1^k)p(\tilde{z}_k|y_1^k)}.$$

We next specialize this formula. First we consider the case $j = R_{k+1} < k$ for which we find

$$E\Big(\sum_1^{k+1}\psi_i|(j,q,l),y_1^{k+1}\Big) = E\Big(\sum_1^k\psi_i|(j,q,l),y_1^k\Big)+\psi((j,q,l),(j,q,l)).$$

For the case $j=k$ we use $q' = q-1(\text{mod } m)$ and get

$$E\Big(\sum_1^{k+1}\psi_i|(k,q,l),y_1^{k+1}\Big)$$
$$= \frac{\sum_{j'<k}\{E(\sum_1^k\psi_i|(j',q',k),y_1^k)+\psi((k,q,l),(j',q',k))\}\gamma_t(j',q',t,l)}{\sum_{j'<k}\gamma_t(j',q',t,l)}.$$

Finally, we find

$$E\left(\sum_{1}^{k+1}\psi_i|y_1^{k+1}\right)=\sum_{k,q,l}p_{k+1}(k,q,l)E\left(\sum_{1}^{k+1}\psi_i|(k,q,l),y_1^{k+1}\right).$$

References

Cavalieri, J., V. E. Eagles, M. Ryan, and K. L. Macmillan (2001). Comparison of four methods for detection of oestrus and diseases in dairy cattle based on time series analysis combined with a kalman filter. *Computers and Electronics in Agriculture 17*, 399–407.

Cavalieri, J., V. E. Eagles, M. Ryan, and K. L. Macmillan (2003). Comparison of four methods for detection of oestrus in dairy cows with resynchronised oestrus cycles. *Aust. Vet. J. 81*, 422–25.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B 39(1)*, 1–38.

Dobson, H., S. L. Walker, M. J. Morris, J. E. Routly, and R. F. Smith (2008). Why is it getting more difficult to successfully ai dairy cows? *Animal in press*.

Elashoff, M. and L. Ryan (2004). An EM algorithm for estimating equations. *J. Comput. Graph. Statist. 13*, 48–65.

Elton, C. and M. Nicholson (1942). The ten-year cycle in numbers of the lynx in canada. *J. Animal Ecol. 11*, 215–244.

Fearnhead, P. and Z. Liu (2007). Online inference for multiple changepoint problems. *J. Roy. Statist. Soc. B 69*, 589–605.

Fearnhead, P. and D. Vasileiou (2008). Bayesian analysis of isochores. *Preprint available from* `www.maths.lancs.ac.uk/~fearnhea/publications`.

Friggens, N. C., M. Bjerring, C. Ridder, S. Højsgaard, and T. Larsen (2008). Improved detection of reproductive status in dairy cows using milk progesterone measurements. *To appear in Reprod. Domestic Anim.*

Fulkerson, W. J., G. J. Sawyer, and I. Crothers (1983). The accuracy of several aids in detecting oestrus in cattle. *Appl. Anim. Ethol. 10*, 199–208.

Heyde, C. and R. Morton (1996). Quasi-likelihood and generalizing the EM-algorithm. *J. Roy. Statist. Soc. B 58*, 317–327.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika 41*, 100–115.

Peters, A. R. and P. J. H. Ball (1995). *Reproduction in cattle*. Blackwell Science, Oxford, UK.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Roelofs, J., F. V. Eerdenburg, W. Hazeleger, N. Soede, and B. Kemp (2006). Relationship between progesterone concentrations in milk and blood and time of ovulation in dairy cattle. *Animal Reproduction Science 91*, 337–343.

Roelofs, J., E. Graat, E. Mullaart, N. Soede, W. Voskamp-Harkema, and B. Kemp (2006). Effects of insemination-ovulation interval on fertilization rates and embryo characteristics in dairy cattle. *Theriogenology 66*, 2173–2181.

Rosen, O., W. Jiang, and M. Tanner (2000). Mixtures of marginal models. *Biometrika 87*, 391–404.

Xu, Z. Z., D. J. McKnight, R. Vishwanath, C. J. Pitt, and L. J. Burton (1998). Estrus detection using radiotelemetry or visual observation and tail painting for dairy cows on pasture. *J. Dairy. Sci. 81*, 2890–2896.