Combining Cattle Activity and Progesterone Measurements Using Hidden Semi-Markov Models

Jared O'CONNELL, Frede Aakmann TØGERSEN, Nicolas C. FRIGGENS, Peter LØVENDAHL, and Søren HØJSGAARD

Hourly pedometer counts and irregularly measured concentration of the hormone progesterone were available for a large number of dairy cattle. A hidden semi-Markov was applied to this bivariate time-series data for the purposes of monitoring the reproductive status of cattle. In particular, the ability to identify oestrus is investigated as this is of great importance to farm management. Progesterone concentration is a more accurate but more expensive method than pedometer counts, and we evaluate the added benefits of a model that includes this variable. The resulting model is biologically sensible, but validation is difficult. We utilize some auxiliary data to demonstrate the model's performance.

Key Words: Dairy cow; EM-algorithm; Oestrus detection; Online data; Streaming data; Time series.

1. INTRODUCTION

Accurate detection of reproductive status in cattle is critical for efficient farm management. A number of on-farm systems for this purpose are in existence. Some of these systems involve either the measuring "activity" (counts from a pedometer per hour) or the concentration of progesterone (measured at the time of milking). Both measures are proxies for the reproductive status of a cow, with progesterone being the more expensive but more accurate predictor. Little work has been done utilizing both measures simultaneously, that is, analyzing them as a multivariate time series. Combining the two measures is expected to lead to improved prediction of reproductive status in addition to furthering the

Jared O'Connell (✉) is DPhil Student, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, England, UK (E-mail: *jared@well.ox.ac.uk*). Frede Aakmann Tøgersen is Statistician, Modeling, Statistics and Risk Analysis, Vestas R&D, Alsvej, Denmark. Nicolas C. Friggens is Head of Research Unit, INRA UMR 791 Modélisation Systémique Appliquée aux Ruminants (MoSAR), AgroParisTech, Paris Cedex, France. Peter Løvendahl is Senior Scientist and Søren Højsgaard is Senior Scientist, Department of Genetics and Biotechnology, Aarhus University, Århus, Denmark.

understanding of the reproductive cycle in cattle. This paper extends some initial ideas by Friggens and Løvendahl (2008). Automated oestrus detection using pedometers has been previously explored using a variety of methodologies, see Firk et al. (2002) for a review.

In this paper we investigate hidden semi-Markov models as a possible methodology for monitoring the reproductive status of cattle based on proxy measurements. Progesterone level is a reliable indicator of oestrus (Friggens et al. 2008), however a financial cost is associated with each measurement taken. Activity measurements are essentially free after the initial set-up overhead, but the data is much noisier leading to missed oestruses and false positives. Univariate models for each measure and a bivariate model using both measures were estimated with the data to assess the utility of each variable and the improvements that can be gained by using both.

In Section 2, we give some biological background and describe the available data. In Section 3, we introduce hidden semi-Markov models and the methods of parameter estimation and prediction and show how they were applied to our data. Section 4 contains results and Section 5 summarizes our work and covers future areas of investigation.

2. BACKGROUND

2.1. THE OVARIAN CYCLE IN CATTLE

Cows are a polyoestrous animal, that is, once started, the oestrus cycle will continue until pregnancy or illness occurs (as opposed to the seasonal oestrus of some animals). The oestrus cycle takes approximately 21 days (but with a large variation) and can be divided into four stages, which can be further grouped into two longer stages (see Table 1). The use of days is an insufficiently granular timescale for the oestrus stage, as oestrus lasts between 6 and 30 hours (Ball and Peters 2004). Ovulation occurs in the day following oestrus, so identifying oestrus allows a farm manager to know when to artificially inseminate a cow.

Following ovulation a structure called the corpus luteum forms in the ovary. The corpus luteum produces the hormone progesterone and remains in the ovary until a few days prior to the next ovulation (the luteal phase), at which time it degenerates rapidly. That is, progesterone is high during the luteal phase and low during the follicular phase. Progesterone directly reflects the biological processes occurring in the ovary and is a very useful indicator of reproductive status. Progesterone can be automatically measured from milk samples, with milking occurring every 6–20 hours (in robotic milking systems). A typical progesterone profile after calving can be viewed in Figure 1 (bottom).

Table 1. Table displaying the ovarian stages of a cow and the related variables. Ovulation occurs within the first day of metoestrus.

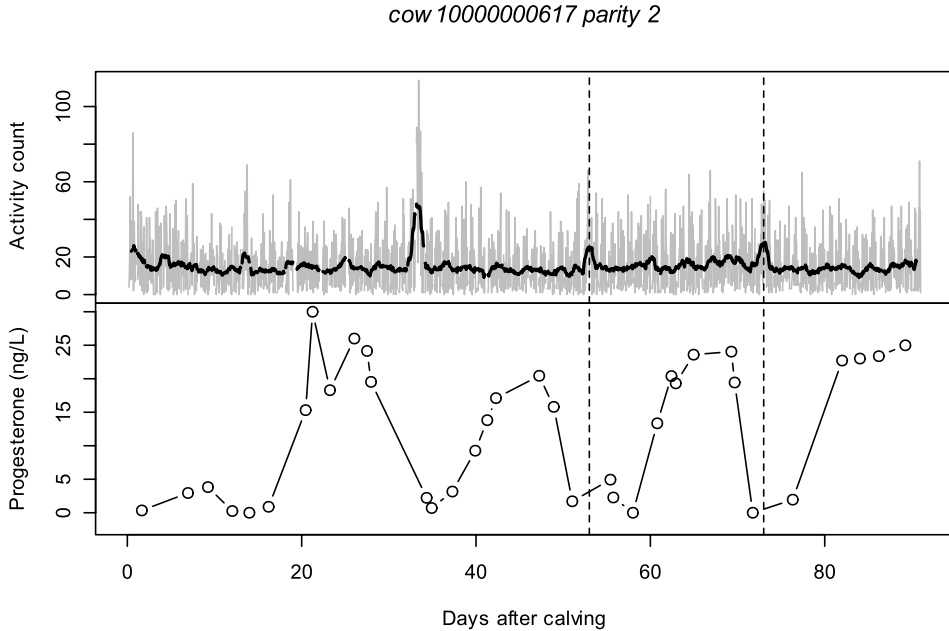| Time	Stage	Metaphase	Activity	Progesterone
Days 0–13	dioestrus	luteal	normal	high
Days 13–15	pro-oestrus		normal	
Day 16	oestrus	follicular	high	low
Days 17–20	metoestrus		normal	

cow 10000000617 parity 2

Figure 1.   Top: The raw hourly activity counts from a pedometer against time since calving (grey), the black line is as 24-hour centred moving average. Bottom: The progesterone concentration over the same time and cow. The dashed vertical lines are times when an artificial insemination occurred (an indicator of oestrus). Note that these occur after a drop in progesterone and during a brief spike in activity levels.

In the period leading up to ovulation the cow will try to attract the attention of a bull by standing to be mounted, mounting other cows or being mounted by other cows. This is sometimes referred to as standing heat and is traditionally how a stockman would identify a cow that is about to ovulate, allowing them to proceed with artificial insemination or bring the cow to a bull. This behaviour leads to an increase in the number of counts on a pedometer the cow is wearing, and so can be exploited for automated detection of oestrus. Having a stockman manually detect oestrus in farms with hundreds of cows is expensive and inaccurate, so automated systems are of great interest. These spikes in activity (together with the corresponding drops in progesterone) are illustrated in Figure 1 (top).

2.2. DATA

The data was collected at the Danish Cattle Research Centre. Activity data consisted of counts per hour from ALPRO$^{®}$ pedometers by DeLaval. Progesterone is measured in nanograms per millilitre (ng/mL) at milking times (Friggens et al. 2008). Progesterone and activity data from 58 cows were analyzed. These cows were selected from a database of 424 cows using filtering rules described below.

Three criteria were applied to individual cows to determine whether they were suitable for this analysis. Firstly, a cow had to have become pregnant for us to consider it. The only sure case of an oestrus occurring is when a cow has become pregnant and we used this in our model validation. Secondly, only cows with both progesterone and activity data

available were used in this work, since we wish to evaluate the benefits of measuring both variables. Thirdly, for reasons described in Section 3.4, series starting at the time of a calving event were desirable, so cows without a single measurement of either variable during the first 21 days after calving were also discarded. After filtering the data, we were left with 58 cows and data spanning 128,578 hours with 112,775 activity counts and 1,917 progesterone measurements. The median time between progesterone measurements was 49 hours.

We rounded the time of progesterone measurements to the nearest hour and merged these with activity measurements to create a time series of bivariate data, suitable to be analyzed as a Markov chain (in respect to the timing of observations). Most hours had only an activity count available, and in these cases progesterone was treated as a missing value. Useful auxiliary information was the date of artificial inseminations and the dates when a cow gave birth. Combining these allowed us to determine when a successful artificial insemination occurred and hence when we can be certain that oestrus occurred.

2.3. ACTIVITY INDEX

Cows are more active during the day and more passive during the night (diurnal cycle) and the level of activity does not remain the same over a cow's lifespan. Hence, we cannot treat activity levels as a stationary process. We choose to remove the effect of this diurnal cycle and any long-term trend. In short, we want to see how active a cow is at a given hour relative to its recent levels of activity. We applied a Holt–Winters model (Holt 1957; Winter 1959) using an implementation by Meyer (2002) to achieve this. Rather than looking at raw activity we worked with $x_t = \log(\text{activity}_t + 1)$, to stabilize variance. A histogram of the transformed variable can be seen in Figure 2 (left).

The Holt–Winters seasonal method is a simple technique for seasonally adjusting data using exponential moving averages. Let x_t be a univariate time series (in our case the transformed activity measure). Forecasting h steps ahead involves variables a_t and s_t with
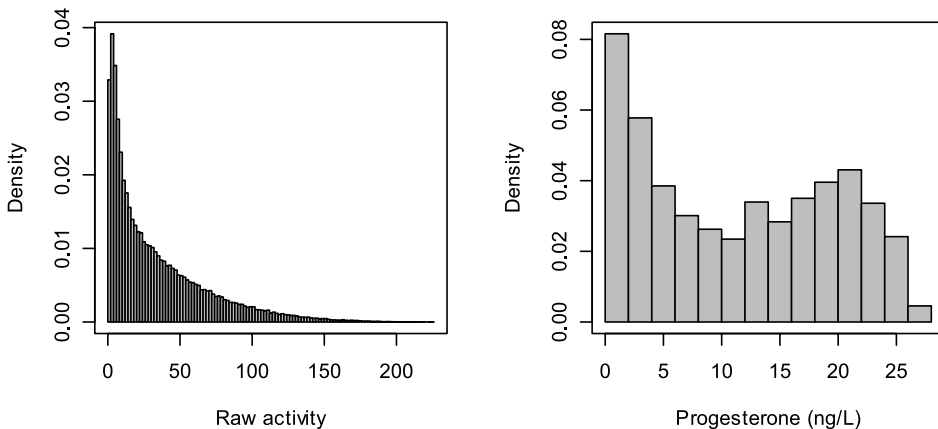

Figure 2.   Left: Histogram of the raw activity data across the entire population. Right: Histogram of the raw progesterone data across the entire population.

the decomposition

$$\hat{x}_{t+h} = \hat{a}_t + \hat{s}_t, \tag{2.1}$$

and the updating formulae

$$\hat{a}_t = \alpha(x_t - \hat{s}_{t-p}) + (1 - \alpha)\hat{a}_{t-1} \qquad \text{and} \qquad \hat{s}_t = \beta(x_t - \hat{a}_t) + (1 - \beta)\hat{s}_{t-p}. \tag{2.2}$$

Here we have a local component, a_t, and a seasonal component, s_t, with period, p, set to 24 to remove the diurnal trend. The parameters α and β are estimated by minimizing the sum of squared errors of one-step-ahead predictions.

If we assume that the Holt–Winters model captures a typical cow behaviour, then the residuals of the one-step-ahead predictions should be large when the cow deviates from normal behaviour. That is, when a cow becomes atypically active (a possible standing heat event) we should observe a large positive residual. Henceforth, we call the one-step-ahead residuals the activity index. A histogram of the activity index can be viewed in Figure 3 (centre). We used the activity index as it does not introduce any lag and does not unnecessarily smooth the data (smoothing is already implicit in the hidden semi-Markov models). Visual inspection of the activity index along with the timing of artificial inseminations and drops in progesterone suggests they were correlated, confirming that this index was a reasonable variable to use (Figure 3, right).

Alternative methods for oestrus detection solely using pedometer readings are in active development, see Jonsson et al. (2008) and references therein. Roelofs et al. (2005) provide a method using simple statistical tests. Firk et al. (2002) review a number of methods. The
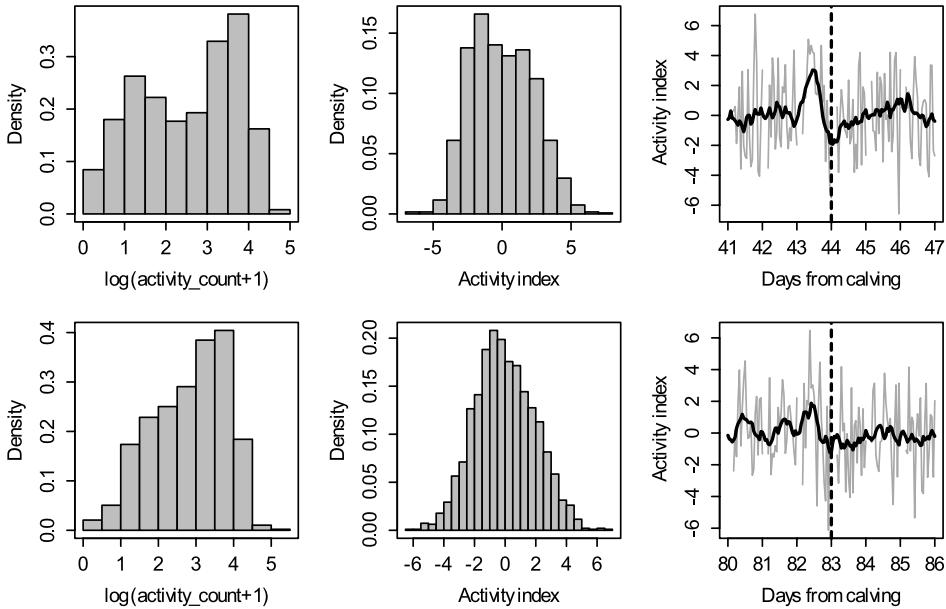

Figure 3.   Left column: Histogram of $\log(activity + 1)$ for two individual cows. Centre: One-step-ahead forecast residuals from a Holt–Winters model for the same cows (the activity index). Right: The activity index in grey with a 24-hour rolling mean overlayed in black, the dashed vertical line is when artificial insemination occurred.

aim of this research was to investigate the relation between progesterone and activity, rather than developing an absolute best technique for oestrus detection with activity readings. Furthermore, different activity measures (or indeed any variable) can be easily "plugged in" to hidden Markov or semi-Markov models at a later date.

3. METHODOLOGY

We give a brief summary of Markov chains, hidden Markov and semi-Markov models, or HMMs and HSMMs respectively. For a complete introduction we refer the reader to Rabiner (1989). Implementation of all models was done using a combination of R (R Development Core Team 2008) and the C programming language. We have released this implementation as an R package called mhsmm (O'Connell and Højsgaard 2009b) and a paper has been submitted on its use (O'Connell and Højsgaard 2009a). We only consider the discrete time case here, as activity measures are hourly counts and the times of progesterone measurements were rounded to the nearest hour.

3.1. DISCRETE MARKOV CHAINS

A discrete Markov chain is a random process (in discrete time) taking discrete values (states) from the state space S, that is, $S_t \in S = \{1, \ldots, J\}$ for $t = 1, 2, \ldots, T$. The process S_t is a Markov chain if it has the Markov property,

$$P(S_{t+1} = s_{t+1} | S_0 = s_0, S_1 = s_1, \ldots, S_t = s_t) = P(S_{t+1} = s_{t+1} | S_t = s_t), \quad (3.1)$$

for any $s_0, s_1, \ldots, s_{t+1} \in S$. That is, the state at any given time $t + 1$ depends only on the previous states through the state at time t. Let $p_{ij} = P(S_{t+1} = j | S_t = i)$ with the properties $\sum_{j=1}^{J} p_{ij} = 1$ and $p_{ij} \geq 0$ so that $\mathbf{P} = (p_{ij})$ is the transition matrix of the Markov chain. To fully specify the model we require the distribution of the initial state $\pi_i = P(S_0 = i)$. For later use we notice that under this model, the probability of spending u consecutive time steps in state i is

$$d_i(u) = P(S_{t+u+1} \neq i, S_{t+u} = i, S_{t+u-1} = i, \ldots, S_{t+2} = i | S_{t+1} = i, S_t \neq i) \quad (3.2)$$

$$= p_{ii}^{u-1}(1 - p_{ii}). \quad (3.3)$$

We call this the duration density and this is inherently geometrically distributed for any Markov chain. This may be an unreasonable assumption for many real-world processes; see Section 3.3 for a discussion.

3.2. HIDDEN MARKOV MODELS

Suppose we can only observe a variable X_t related to the state S_t but not the state itself. The conditional distribution of the observed variable X_t conditioned on the unobserved (or hidden) state S_t is referred to as the emission distribution. We refer to the parameters defining such a process as a hidden Markov model, henceforth referred to as an HMM.

These models have been used for a variety of different applications, such as speech recognition (Rabiner 1989), weather modelling (Hughes, Guttorp, and Charles 1999) and DNA sequence analysis (Krogh, Mian, and Haussler 1994).

In addition to the usual Markov chain parameters, π and $\mathbf{P}$, an HMM also requires an emission distribution to be defined:

$$b_i(x_t) = P(X_t = x_t | S_t = i). \tag{3.4}$$

For example $b_i(x)$ may be a multivariate Gaussian distribution. We assume that the emitted variables are conditionally independent given the underlying state, that is,

$$P(X_t = x_t, X_{t-1} = x_{t-1} | S_t = i) = P(X_t = x_t | S_t = i). \tag{3.5}$$

A full HMM is hence specified by $\theta = (\pi, \mathbf{P}, b)$.

The Baum–Welch algorithm is the original procedure for estimating the parameters of a HMM (Baum et al. 1970). This technique was later grouped with a more general class of algorithms for incomplete data, named the Expectation–Maximization (EM) algorithm (Dempster, Laird, and Rubin 1977). Rabiner (1989) provides a very clear overview.

3.3. HIDDEN SEMI-MARKOV MODELS

A severe limitation of standard HMMs is the implicit geometrically distributed duration time of states in the model (Equation (3.3)), as many of the real-world problems we wish to model do not have geometric sojourn times. For example, the reproductive states of cows are not a memoryless process since a follicular stage is likely to occur after 18 days in a luteal stage. A potential solution to this issue is to explicitly estimate $d(u)$, producing what is referred as a hidden semi-Markov model, henceforth called an HSMM. Ferguson (1980) was the first to propose such models along with an algorithm to fit them; Rabiner (1989) provides a good summary. Guédon (2003) developed a more efficient algorithm and a method to deal with right censoring which we have implemented.

In a standard discrete Markov chain the state duration density has the form $d_i(u) = p_{ii}^u(1 - p_{ii})$, whereas in a semi-Markov model, we model $d_i(u)$ explicitly. That is, in addition to the parameters already defined for the HMM, we now have explicit probabilities for our state duration density, $\theta = (\pi, \mathbf{P}, b, d)$. Guédon (2003) utilizes a non-parametric probability mass function of $d(u)$ in his derivations, but then proposes an ad hoc method to introduce parametric distributions (for example, Poisson). The latter procedure has been followed here.

3.3.1. Parameter Estimation

The complete data likelihood of a HSMM is

$$P(X = x, S = s; \theta) = \pi_{s_1^*} d_{s_1^*}(u_1) \left\{ \prod_{r=2}^{R} p_{s_{r-1}^* s_r^*} d_{s_r^*}(u_r) \right\} p_{s_{R-1}^* s_R^*} D_{s_R^*}(u_R) \prod_{t=1}^{T} b_{s_t}(x_t) \tag{3.6}$$

where s_r^* is the rth visited state and u_r is the time spent in that state. Guédon proposed using the survivor function

$$D_i(u) = \sum_{v \geq u} d_i(v) \tag{3.7}$$

for the sojourn time in the last state, so we do not have to assume the process is leaving a state immediately after time T. This is particularly important for online applications where we wish to estimate the most recent state so we can monitor some process.

As we have not observed the state sequence, maximizing this likelihood constitutes an incomplete data problem. A local maximum can be found using the Expectation–Maximization, or EM, algorithm. We briefly outline the procedure below.

The EM algorithm involves iterating over two steps until convergence. In the E-step, we calculate the expected complete data likelihood given the value of the parameters at iteration k and the observed data:

$$Q\big(\theta\big|\theta^{(k)}\big) = E\big[\log\big(P(X = x, S = s; \theta)\big)\big|X = x; \theta^{(k)}\big]. \tag{3.8}$$

This term is typically broken down into a sum of terms involving subsets of the parameters. The M-step then involves choosing $\theta^{(k+1)}$ as the values that maximize $Q(\theta|\theta^{(k)})$. These steps are repeated until convergence. We now provide the specific terms calculated for an HSMM.

The E-step For HSMMs, the E-step involves estimating three separate terms: the probability of being in state i at time t,

$$\gamma_t(i) = P(S_t = i|X = x; \theta), \tag{3.9}$$

the probability that the process left state i at time t and entered state j at $t + 1$,

$$\xi_t(i, j) = P(S_t = i, S_{t+1} = j|X = x; \theta), \tag{3.10}$$

and finally the expected number of times a process spent u time steps in state i,

$$\eta_{i,u} = P(S_u \neq i, S_{u-v} = i, v = 1, \ldots, u|X = x; \theta)$$
$$+ \sum_{t=1}^{T} P(S_{t+u+1} \neq i, S_{t+u-v} = i, v = 0, \ldots, u - 1, S_t \neq i|X = x; \theta). \tag{3.11}$$

These values can be calculated via a dynamic programming method known as the forward-backward algorithm. It should be noted that the derivations and algorithms involved are complex; we refer the reader to Guédon (2003) for details.

The M-step The starting and transition probabilities are implicit from the definition of the Markov chain,

$$\hat{\pi}_i' = \gamma_1(i), \qquad \hat{p}_{ij}' = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \sum_{i \neq j} \xi_t(i, j)}. \tag{3.12}$$

Guédon provides derivations for $d_i(u)$ as a non-parametric probability mass function, but then proposes an ad hoc solution for using parametric distributions with $\eta_{i,u}$ as weights

which we have followed. We assume that the sojourn times are Gamma distributed, that is, $U_r|S_r = i \sim \Gamma(c_i, \lambda_i)$. We estimated the parameters as follows: the likelihood for the Gamma distribution can be maximized with respect to its parameters by solving

$$\log(\hat{c}_i) - \psi(\hat{c}_i) = \log(\bar{u}_i) - \overline{\log u_i} \tag{3.13}$$

where $\psi()$ is the digamma function. We use

$$\bar{u}_i = \frac{\sum_u \eta_{i,u} u}{\sum_u \eta_{i,u}}, \qquad \overline{\log u_i} = \frac{\sum_u \eta_{i,u} \log(u)}{\sum_u \eta_{i,u}} \tag{3.14}$$

and then solve the equation using Newton's method (Choi and Wette 1969). The scale parameter is estimated as $\hat{\lambda}_i = \hat{c}_i / \bar{u}_i$.

We assume X_t are normally distributed given $S = i$, that is, $X_t|S_t = i \sim N(\mu_i, \sigma_i^2)$. We can re-estimate the Gaussian parameters via the equations:

$$\hat{\mu}_i = \frac{\sum_{t=1}^T \gamma_t(i) x_t}{\sum_{t=1}^T \gamma_t(i)}, \qquad \hat{\sigma}_i^2 = \frac{\sum_{t=1}^T \gamma_t(i)(x_t - \hat{\mu}_i)^2}{\sum_{t=1}^T \gamma_t(i)}. \tag{3.15}$$

Sansom and Thomson (2001) provide derivations for these re-estimating equations. In the case where X_t is multivariate, for our application we assumed that the individual components of X_i were conditionally independent given the current state and so take the joint density as the product of the individual densities.

3.3.2. Prediction: The Viterbi Algorithm

Given a model, we are interested in the most likely sequence of states, given the sequence of observations. That is, we wish to find the sequence of states that maximizes $P(S|X, \theta)$. Calculating $P(S|X, \theta)$ for every possible state sequence is not computationally feasible. A dynamic programming technique known as the Viterbi algorithm (Forney 1973), can be used to maximize $P(S|X, \theta)$ in a computationally feasible manner. It should be noted that this is distinct from taking $\mathrm{argmax}_i \gamma_t(i)$ as the state at time t, which is the individually most likely state. We start by defining the variable

$$\delta_t(j) = \max_{s_1,\dots,s_{t-1}} P(S_{t+1} \neq j, S_t = j, S_1 = s_1, \dots, S_{t-1} = s_{t-1}, X_1 = x_1, \dots,$$
$$X_{t-1} = x_{t-1}; \theta) \tag{3.16}$$

with the special case at the last time step

$$\delta_T(j) = \max_{s_1,\dots,s_{T-1}} P(S_T = j, S_1 = s_1, \dots, S_{T-1} = s_{T-1}, X_1 = x_1, \dots, X_T = x_T; \theta), \tag{3.17}$$

where we are only concerned with the probability of being in state j at time T, not with leaving the state at $T + 1$. The variable δ can be calculated via induction with the initialization step

$$\delta_1(j) = b_j(x_1) d_j(1) \pi_j, \tag{3.18}$$

the recursion step

$$\delta_t(j) = b_j(x_t) \max \left\{ \max_{1 \leq u \leq t} \left[\left(\prod_{v=1}^{u-1} b_j(x_{t-v}) \right) d_j(u) \max_{i \neq j} \left(p_{ij} \delta_{t-u}(i) \right) \right]; \right.$$
$$\left. \left(\prod_{v=1}^{t} b_j(x_{t-v}) \right) d_j(t+1)\pi_i \right\}, \tag{3.19}$$

and the terminating step

$$\delta_T(j) = b_j(x_T) \max \left\{ \max_{1 \leq u \leq T} \left[\left(\prod_{v=1}^{u-1} b_j(x_{T-v}) \right) D_j(u) \max_{i \neq j} \left(p_{ij} \delta_{T-u}(i) \right) \right]; \right.$$
$$\left. \left(\prod_{v=1}^{T} b_j(x_{T-v}) \right) D_j(T)\pi_i \right\}. \tag{3.20}$$

The likelihood of the optimum state sequence is then just $\max_i \delta_T(i)$. To obtain the optimal sequence of states we must retain the values of u and i that maximize $\delta_t(j)$ at each step. We record the most likely value of u, that is the most likely sojourn time,

$$\upsilon_t(i) = t - \operatorname*{argmax}_{u,t} \left\{ \max_{1 \leq u \leq t} \left[\left(\prod_{v=1}^{u-1} b_i(x_{t-v}) \right) d_j(u) \max_{i \neq j} \left(p_{ij} \delta_{t-u}(j) \right) \right]; \right.$$
$$\left. \left(\prod_{v=1}^{t} b_i(x_{t-v}) \right) d_j(t+1)\pi_j \right\}, \tag{3.21}$$

and the most likely preceeding state,

$$\psi_t(j) = \operatorname*{argmax}_{i \neq j} \left[p_{ij} \delta_{\upsilon_t(i)}(j) \right], \qquad 2 \leq t \leq T. \tag{3.22}$$

Finally, we apply a backtracking algorithm to find the optimal state sequence:

1. Set $t = T$

2. $s'_t = \operatorname{argmax}_i(\delta_t(i))$

3. $s'_{\upsilon_t(s'_t)+1}, \ldots, s'_{t-1} = s'_t$

4. $s'_{\upsilon_t(s'_t)} = \psi_t(s'_t)$

5. Set $t = \upsilon_t(s'_t)$

6. If $t = 1$ stop, else return to 3

3.3.3. Missing Data

Missing data occurs in both of the observed variables for our application. Pedometer readings for a given hour may fail to be recorded due to hardware problems or a cow being away from the stable. There are many missing progesterone measurements due to the fact it is measured at a much sparser resolution than activity (less than 2% of activity

recordings have a matching progesterone measurement). As such, we require a method for handling missing observations, which is straightforward for HSMMs. If an observation x_t is missing, we set $b_j(x_t) = 1$ for all j, that is, there is no contribution to the likelihood at that time step.

3.4. APPLICATION TO CATTLE DATA

Three models were investigated: two univariate models using each variable and one model using both variables. The states identifiable by each combination of variables differ. For each univariate model we used three states, but with different definitions. The activity-only model contained states

$$S^{\text{Activity}} = \{\textit{post-partum anoestrus}, \textit{standing heat}, \textit{not standing heat}\}$$

(Figure 4, top). For progesterone measurements we had

$$S^{\text{Progesterone}} = \{\textit{post-partum anoestrus}, \textit{luteal}, \textit{follicular}\}$$

(Figure 4, bottom). Standing heat occurs near the beginning of the follicular state but the follicular state is much longer. The bivariate model used six states, the combination of activity and progesterone allowing us to break the follicular stage up into three different sub-stages in addition to the PPA and luteal states (Figure 5).

We define $t = 1$ as the hour immediately following a calving. This is a useful way to arrange the data as we know that the cow will be in post-partum anoestrus at this time. Gaussian emission distributions were used for each variable. We worked with
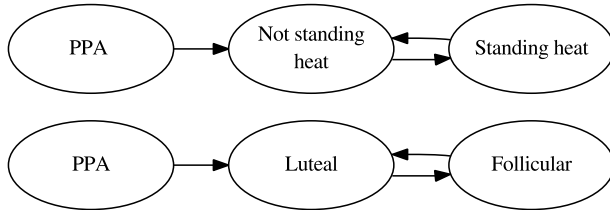

Figure 4.   Top: The underlying states for a univariate activity model. Bottom: The underlying states for a univariate progesterone model.
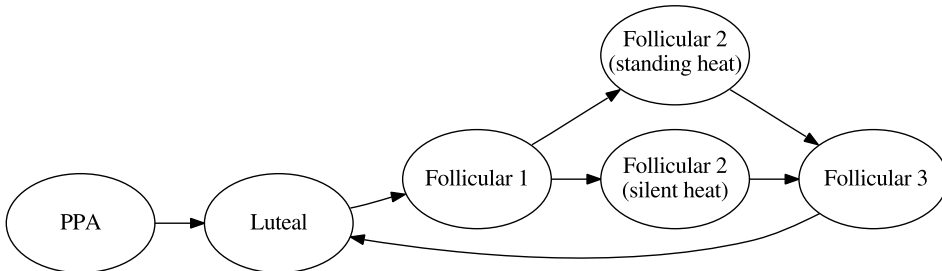

Figure 5.   The hidden states used for the bivariate HSMM. The use of both variables allows us to break the follicular stage up into three different sub-stages.

log(progesterone + 1) rather than raw progesterone to stabilize the variance. Gamma distributions were used for the distribution of sojourn times for each state. Starting values for distribution parameters were found via manual inspection of the data and biological knowledge of the ovarian cycle as detailed in Section 2.1.

4. RESULTS

Validation for this data is difficult as the true timing of ovulation is never known. One proxy is when a successful artificial insemination occurs, that is, an insemination that results in a cow giving birth approximately 280 days later. We take two approaches here, first investigating the ability of the activity-only model to predict the states of a progesterone-only model. Secondly, we validate the bivariate model using successful insemination times as proxies for ovulation.

4.1. PREDICTION OF FOLLICULAR STATE BY ACTIVITY-ONLY MODEL

We estimated parameters for two separate univariate models for each variable using the methods described in Section 3. The Viterbi algorithm was applied to find the optimal state sequences. We then looked at how well the state sequence from the activity model aligned with that from the progesterone model, treating the state sequence from progesterone as validation data. According to biological knowledge, standing heat should occur near the beginning of the follicular stage (Ball and Peters 2004). We considered a predicted *standing heat* from the activity model that was bounded (in time) by a predicted *follicular* state from the progesterone model to be a correct prediction. We do not claim that the state sequence from progesterone is true, but a closer reconstruction of the biological states of the cow.

Two possible prediction errors can occur in this framework; false positive (prediction of standing heat outside a follicular state) and false negative (failure to predict standing heat within a follicular stage). An example of some correctly predicted follicular stages is shown in Figure 6. The results are summarized in Table 2.

4.2. OESTRUS DETECTION

We compared the bivariate model and the activity-only univariate model as online systems by using information about successful artificial inseminations and applying cross-validation. Cross-validation was used as follows. We iterated through a set of 58 cows with successful pregnancies, excluding one cow's full data sequence and estimating the HSMM parameters using the remaining data. We then applied the Viterbi algorithm to the left-out cows' sequence up until the time of successful insemination. Finally, we compared the time of the most recent predicted (if any) oestrus event by the models to the artificial insemination event. Artificial insemination is thought to be best applied within 24 hours following oestrus so we expect a positive difference between AI and the last detected oestrus hour.

A kernel density estimate of the bivariate models time differences can be viewed in Figure 7 (left) and a comparison of the two models in Figure 7 (right). The results for each model are summarized in Table 3 where we can see that the bivariate model had slightly
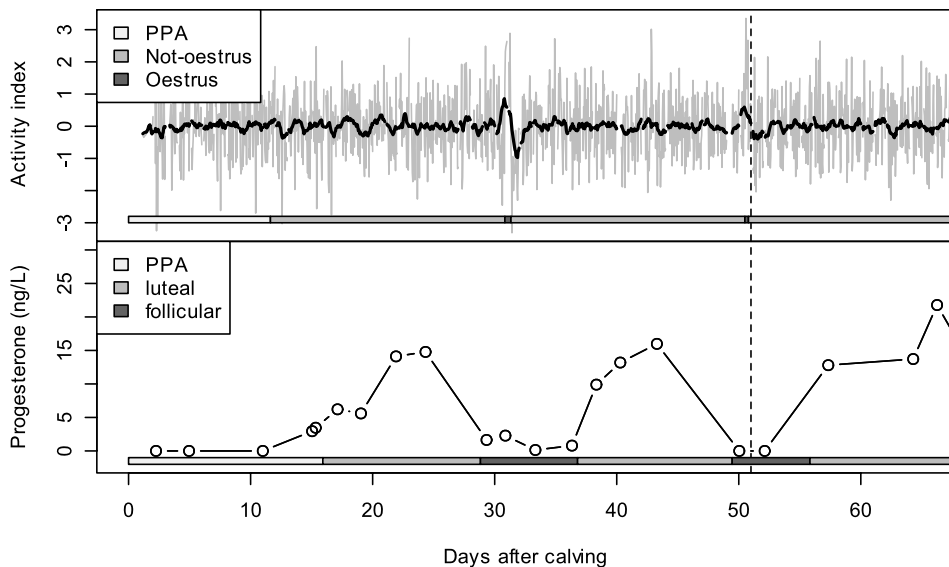
Figure 6. Top: The activity index against time since calving. Bottom: The progesterone concentration over the same time and cow. Predicted states by the respective univariate models are shown on the horizontal shaded bars. This is an example where an activity-only model has correctly found oestrus events occur within a follicular stage.

Table 2. Success and error rates for prediction of a follicular stage by an univariate activity HSMM. TP = True positive, FP = False positive, FN = False negative.

	Sensitivity $\frac{TP}{TP+FN}$	Error rate $\frac{FP}{TP+FP}$
Count	97/137	16/113
Proportion	0.71	0.14

lower times indicating better performance. The activity-only model failed to detect any oestrus in 7 of the 58 cases as well as the two obvious false negatives that can be seen in the graph, a false negative rate of at least 0.16. The bivariate model essentially uses the activity measure to narrow the window of oestrus with the follicular stage, so it is not surprising the models are comparable in terms of oestrus timing (when the activity model actually finds the oestrus).

We have limited our analysis to insemination events that led to a pregnancy. The majority of inseminations do not lead to a pregnancy and we have excluded these failed inseminations from the analysis. Inseminations may fail due to incorrect timing, technical problems with the process or cow infertility. We can only be certain an oestrus has occurred if a cow becomes pregnant and hence have used these events to validate our model.
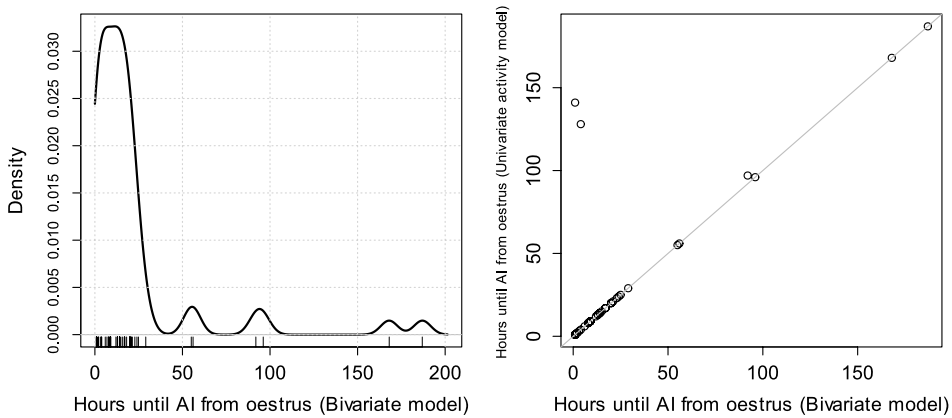
Figure 7. Left: Kernel density estimate of the time difference between successful AI and the most recently detected oestrus state by the bivariate model. Right: The univariate model against the bivariate model for time to successful AI from detected oestrus. Note the two large values from the univariate model, certainly false negatives.

Table 3. Cumulative table of the time between a predicted oestrus event and an actual successful artificial insemination for the bivariate and univariate (activity-only) models. Predictions that occurred less than one day before the actual insemination can be considered successful, predictions between one and two days are borderline. The bivariate model predicts more oestruses within one to two days of an insemination.

Days	1	2	3	4	5	6	7	8
Time to AI < Days (bivariate)	0.84	0.89	0.93	0.95	0.96	0.96	0.96	1.00
(activity only)	0.68	0.73	0.77	0.77	0.80	0.84	0.84	0.88

5. DISCUSSION

We have investigated and compared univariate and bivariate HSMMs for two different variables related to cattle reproduction. Approximately 70% of follicular states were predicted by the activity-only model. This is less accurate than results reported by Roelofs et al. (2005). However those detections rates were for behavioral oestrus events, whereas our data also contains silent oestruses. The false positive rate of 14% is in line with their and other results (Jonsson et al. 2008). The results from our bivariate model indicate that the introduction of progesterone can reduce the number of false positives. The bivariate model was slightly better in terms of predicting the timing of oestrus. In 9 out of 58 cases the activity-only model failed to detect the relevant oestrus.

Progesterone has a financial cost associated with each measurement taken due to the chemicals used, while activity recordings are essentially free after initial overhead. Whether the reduction in false positives and negatives and slightly improved timing justifies these costs is a matter for further investigation. It would also be useful to investigate how the frequency of progesterone measurements affects accuracy. Another possibility is using activity as an indicator as to when to begin progesterone monitoring. For example, progesterone measurements are not useful during post-partum anoestrus. We found an

activity-only model could often identify when a cow has left post-partum anoestrus (although we have not formally reported the results here); this could be a trigger to begin progesterone monitoring.

ACKNOWLEDGEMENTS

REFERENCES

Ball, P. J. H., and Peters, A. R. (2004), *Reproduction in Cattle* (3rd ed.), Oxford: Blackwell Publishing.

Baum, L., Petrie, T., Soules, G., and Weiss, N. (1970), "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annals of Mathematical Statistics*, 41, 164–171.

Choi, S., and Wette, R. (1969), "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias," *Technometrics*, 11, 683–690.

Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1–38.

Ferguson, J. (1980), Hidden Markov Analysis: An Introduction, *Hidden Markov Models for Speech*.

Firk, R., Stamer, E., Junge, W., and Krieter, J. (2002), "Automation of Oestrus Detection in Dairy Cows: A Review," *Livestock Production Science*, 75, 219–232.

Forney, G. Jr. (1973), "The Viterbi Algorithm," *Proceedings of the IEEE*, 61, 268–278.

Friggens, N. C., and Løvendahl, P. (2008), "The Potential of On-Farm Fertility Profiles: In-Line Progesterone and Activity Measurements," in *Fertility in Dairy Cows: Bridging the Gaps*, *British Society of Animal Science*, eds. N. C. Friggens, M. D. Royal, R. Smith, Cambridge University Press, Cambridge, pp. 72–78.

Friggens, N. C., Bjerring, M., Ridder, C., Højsgaard, S., and Larsen, T. (2008), "Improved Detection of Reproductive Status in Dairy Cows Using Milk Progesterone Measurements," *Reproduction in Domestic Animals*, 43, 113–121.

Guédon, Y. (2003), "Estimating Hidden Semi-Markov Chains From Discrete Sequences," *Journal of Computational & Graphical Statistics*, 12, 604–639.

Holt, C. (1957), "Forecasting Trends and Seasonals by Exponentially Weighted Moving Averages," *ONR Memorandum*, 52.

Hughes, J., Guttorp, P., and Charles, S. (1999), "A Non-Homogeneous Hidden Markov Model for Precipitation Occurrence," *Journal of the Royal Statistical Society (Series C): Applied Statistics*, 48, 15–30.

Jonsson, R., Bjorgvinsson, T., Blanke, M., Poulsen, N., Højsgaard, S., and Munksgaard, L. (2008), "Oestrus Detection in Dairy Cows Using Likelihood Ratio Tests," *The International Federation of Automatic Control*, 658–663.

Krogh, A., Mian, I., and Haussler, D. (1994), "A Hidden Markov Model that Finds Genes in E. Coli DNA," *Nucleic Acids Research*, 22, 4768–4778.

Meyer, D. (2002), "Naive Time Series Forecasting Methods," *R News*, 2, 7–10.

O'Connell, J., and Højsgaard, S. (2009a, submitted), "Hidden Semi Markov Models for Multiple Observation Sequences—The MHSMM Package for R," *Journal of Statistical Software*.

O'Connell, J., and Højsgaard, S. (2009b), "MHSMM: Parameter Estimation and Prediction for Hidden Markov and Semi-Markov Models for Data with Multiple Observation Sequences," *http://cran.r-project.org/web/packages/mhsmm/index.html*. R package version 0.3.0.

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

Rabiner, L. (1989), "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, 77, 257–286.

Roelofs, J., van Eerdenburg, F., Soede, N., and Kemp, B. (2005), "Various Behavioral Signs of Estrous and their Relationship with Time of Ovulation in Dairy Cattle," *Theriogenology*, 63, 1366–1377.

Sansom, J., and Thomson, P. (2001), "Fitting Hidden Semi–Markov Models to Breakpoint Rainfall Data," *Journal of Applied Probability*, 38, 142–157.

Winter, P. (1959), "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, 6, 324–342.