Original article

A periodic analysis of longitudinal binary responses: a case study of clinical mastitis in Norwegian Red cows

Yu-mei Chang[a*], Daniel Gianola[a,b], Bjørg Heringstad[b,c], Gunnar Klemetsdal[b]

[a] Department of Dairy Science, University of Wisconsin, Madison, USA
[b] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences
[c] Geno Breeding and A. I. Association, Norway

Abstract – A Bayesian procedure for analyzing longitudinal binary responses using a periodic cosine function was developed. It was assumed that, after adjustment for "seasonal" effects, the oscillation of the underlying latent variables for longitudinal binary responses was a stationary series. Based on this assumption, a single dimension sinusoidal analysis of longitudinal binary responses using the Gibbs sampling and Metropolis algorithms was implemented in a study of clinical mastitis records of Norwegian Red cows taken over five lactations.

clinical mastitis / periodic analysis / sinusoidal analysis / threshold model

1. INTRODUCTION

The study of stationary time series by means of auto-covariance functions is often referred to as a "time domain" analysis. This can be shown to be equivalent to a "frequency domain" analysis, which is based on a spectral representation of a time series [2]. In the latter, characteristics of a process are summarized in terms of frequency, amplitude and phase. Frequency is the number of complete cycles of an oscillation occurring within a period of time; amplitude is the maximum value of a periodically varying series, and phase is the distance between the origin and the nearest peak in time [2].

The analysis of co-movement between time series is of interest in economics, and different approaches have been developed for testing whether several economic variables have common trends or common cycles [9, 10, 16]. However, a detectable periodic component must be present in each time series

---

* Corresponding author: y.chang@leeds.ac.uk

before these tests are applicable. Stoffer [15] proposed a random effects approach for real-valued time series, in which frequency was partitioned into a frequency common to all sequences, plus a random frequency effect peculiar to an individual series. This can enhance a standard frequency analysis, because, rather than focusing on an individual series only, use is made of the relationships among all time series. If an individual series contains the frequency of interest but with small amplitude, or with a lot of noise, information on other series containing the similar harmonic curve may help to capture a particular feature of the process [15].

Mastitis is an inflammation of the mammary gland associated with bacterial infection; it is the most costly disease in dairy cattle [6]. A periodic analysis implies that a signal repeats itself periodically, and this seems to be so for clinical mastitis in cows when followed over several lactations [7]. In statistical analysis, presence or absence of infection is typically related to some latent variable, called "liability". The trajectory of the liability to clinical mastitis for each animal could be represented as the sum of some systematic effects, plus an oscillation of a cosine function at certain frequency, plus white noise (random residual). The frequency and amplitude of the oscillation provide information about the variation of liabilities over time.

Figure 1(a) gives examples of cosine functions, $Y_t = a \cos\{2\pi f(t - \theta)\}$, with different amplitudes ($a = 0.3$ *versus* $a = 0.6$) at the same frequency ($f = 0.2$) and phase ($\theta = 1$). For example, if a cow had a larger amplitude, it would be more susceptible to clinical mastitis at certain periods of lactation than a cow with a lower amplitude. In Figure 1(b), two cows have the same amplitude ($a = 0.3$) and phase ($\theta = 1$), but different frequencies ($f = 0.2$ *versus* $f = 0.5$). The cow with higher frequency (dotted curve) would be expected to contract mastitis more frequently. The phase parameter θ indicates how early a first mastitis case would occur. As shown in Figure 1(c), the cow associated with the solid curve has highest risk of first mastitis at period 1, whereas the other cow is more susceptible to a first mastitis case at period 3 of the oscillations. On this basis, animal breeders would select cows that, genetically, have smaller amplitude and frequency, and larger phase of the cycle.

If a random effect is included in a sinusoidal model, each cow would have a specific frequency, amplitude and phase. In the sense of the time domain, this means that each cow would have different mastitis risks at different time points. For instance, the amplitude of the signal might be modeled with fixed and random components, and a larger absolute value of the amplitude would indicate
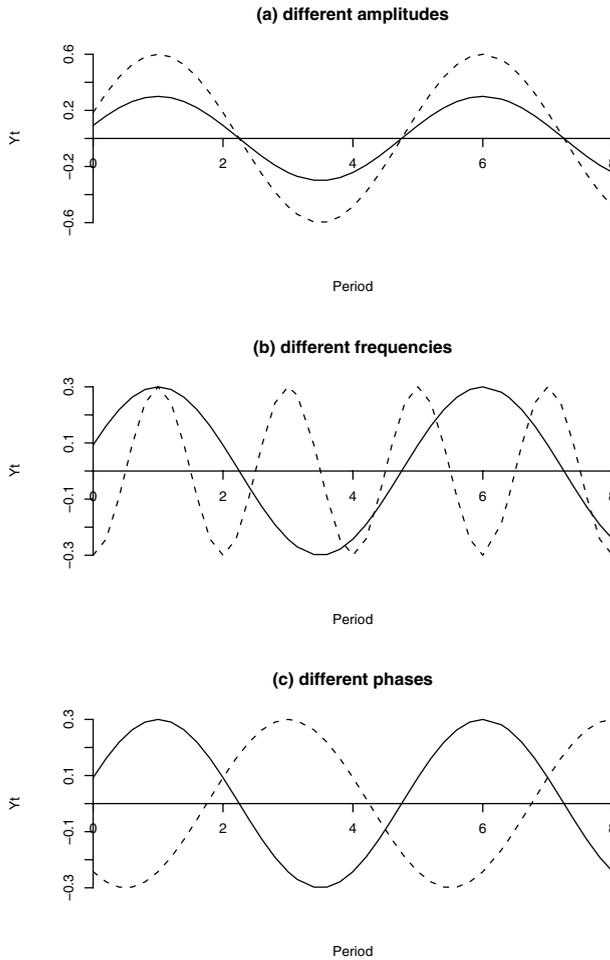
(a) different amplitudes

(b) different frequencies

(c) different phases

Figure 1. Examples of periodical time series of a cosine function at (a) different amplitudes, (b) different frequencies and (c) different phases.

a higher risk of mastitis. If two cows have the same frequency parameter, the one with a larger amplitude would have a higher chance of developing mastitis.

Our objective is to embed Stoffer's [15] approach into a generalized nonlinear mixed effects model for binary response variables. Latent variables representing, *e.g.*, "liability to disease", will be modelled such that each subject has specific frequency and amplitude parameters over time. A Bayesian model with Markov chain Monte Carlo computations is presented in Section 2 and an application to clinical mastitis data in dairy cattle is given.

2. MATERIALS AND METHODS

2.1. Statistical model

The classical threshold model of quantitative genetics was adopted [4, 5, 17]. For a binary response (the idea can be extended to ordered polychotomous data), this model postulates the existence of an unobserved underlying or latent variable l (*e.g.*, liability to disease) and a fixed conceptual threshold such that, if liability exceeds the threshold, then disease occurs. In a binary response situation, the threshold can be set arbitrarily to zero. Consider a longitudinal case, where the binary response variable y_{it} for cow i at period t takes the value 1 if the cow has the disease, and 0 otherwise, according to the rule:

$$y_{it} = \begin{cases} 1, \text{ if } l_{it} \geq 0 \\ 0, \text{ if } l_{it} < 0, \end{cases} \tag{1}$$

where l_{it} is the liability underlying y_{it}.

Liability was modeled as a function of some systematic effects and of a cosine function of amplitude, frequency and phase, plus random noise. The model for liability was

$$l_{hijkt} = \mathbf{x}'_{hijkt}\boldsymbol{\beta} + \mathbf{z}'_{h,hijkt}\mathbf{h} + \mathbf{z}'_{p,hijkt}\mathbf{p} + a_{j,k}\cos\{2\pi f_{j,k}(t-\theta)\} + e_{hijkt}, \tag{2}$$

where l_{hijkt} is the liability of cow i at period t, calving in herd h, age-season j and daughter of sire k. Here, $\mathbf{x}'_{hijkt}$ is an incidence vector relating liability to some systematic effects contained in $\boldsymbol{\beta}$; $\mathbf{z}'_{h,hijkt}$ is an incidence vector relating liability to herd effects $\mathbf{h}$; $\mathbf{z}'_{p,hijkt}$ is an incidence vector relating liability to cow-specific effects contained in the vector $\mathbf{p}$. Further, $a_{j,k}$ and $f_{j,k}$ are the amplitude (the maximum or minimum value of the cosine series) and frequency (the number of complete cycles of the cosine oscillation within a period of time), respectively, of a cosine function peculiar to age-season of first calving j and sire family k. θ is the distance between the origin and the first peak (phase), assumed to be a known constant across all series; this can be relaxed if different phases are evident across series. Finally, $e_{hijkt} \sim \text{NIID}(0, 1)$ where NIID stands for "normally, independently and identically distributed."

The assumption is that if liability is "adjusted" for the effects of $\boldsymbol{\beta}$, $\mathbf{h}$ and $\mathbf{p}$, the remaining part of the model produces a stationary series. The oscillation of the cosine curve depends on a "fixed" part (*i.e.*, the age-season of first calving j) and a random part (*i.e.*, the sire family k). The term e_{hijkt} is a random noise, assumed to follow a normal process with mean 0 and variance 1. Since liability cannot be observed, the residual standard deviation is taken as unit of

measurement. The conditional distribution of the liability of cow i at time t is then

$$l_{hijkt}|\boldsymbol{\beta}, \mathbf{h}, \mathbf{p}, a_{j,k}, f_{j,k}, \theta, y_{hijkt}$$
$$\sim \text{TN}(\mathbf{x}'_{hijkt}\boldsymbol{\beta} + \mathbf{z}'_{h,hijkt}\mathbf{h} + \mathbf{z}'_{p,hijkt}\mathbf{p} + a_{j,k}\cos\{2\pi f_{j,k}(t-\theta)\}, 1),$$

where TN stands for a normal distribution truncated following the rule in (1). That is, the liability has to be larger or equal to zero if a cow had mastitis and smaller than zero, otherwise.

The amplitude and frequency parameters were modeled hierarchically as

$$\begin{bmatrix} a_{j,k} \\ f_{j,k} \end{bmatrix} = \begin{bmatrix} b_j \\ \omega_j \end{bmatrix} + \begin{bmatrix} s_{1,k} \\ s_{2,k} \end{bmatrix}.$$

Here, b_j is the "fixed" amplitude effect, peculiar to age-season of first calving class j, and ω_j is a frequency that is peculiar to all cows in age-season of first calving class j. Further, $s_{1,k}$ and $s_{2,k}$ are random amplitude and frequency effects, respectively, which are peculiar to all cows that are daughters of sire k. For simplicity, we will assume that the frequency ω_j is a known constant across all series, and drop the subscript for age-season class from the notation. The model for liability is nonlinear with respect to frequency, but linear related to amplitude.

Bayesian MCMC methods, such as Gibbs sampling and the Metropolis algorithm, were used for drawing samples from posterior distributions of interest. The sampling space of each element of $\boldsymbol{\beta}$, $\mathbf{p}$ and $\mathbf{h}$ is the real line. The parameter space for amplitude, $a_{j,k} = b_j + s_{1,k}$, is also the real line. At least two time units are required for a cycle to be finished, so the "fastest" curve (that having the highest frequency) that can be observed is one with a period of two time units. On the other hand, the "slowest" cosine curve is one with a period of T time units, where T is the maximum number of time units. Therefore, the frequency $f_k = \omega + s_{2,k}$ lies between $\frac{1}{T}$ (slowest curve) and $\frac{1}{2}$ (fastest curve).

2.2. Prior and posterior distributions

A bounded uniform prior distribution $U[\beta_{\min}, \beta_{\max}]$ was assigned to each of the elements of $\boldsymbol{\beta}$, where $\beta_{\min} = -99$ and $\beta_{\max} = 99$. The vector of herd effects $\mathbf{h}$ was assumed to follow the normal process

$$\mathbf{h} \sim \text{N}(0, \mathbf{I}\sigma_h^2),$$

where σ_h^2 is the variance between herds, and the cow-specific effects $\mathbf{p}$ (usually termed "permanent environmental effects" in animal breeding) were assumed to be normally distributed, *a priori*, as

$$\mathbf{p} \sim \mathrm{N}(0, \mathbf{I}\sigma_p^2),$$

where σ_p^2 is the variance of the distribution.

Independent bounded uniform priors were assigned to the "fixed" amplitude, such that $b_j \sim \mathrm{U}[0, b_{\max}]$ for all j, where $b_{\max} = 9$. The random sire variables $s_{1,k}$ and $s_{2,k}$, interpretable as half of the additive genetic values of a sire, were assumed to follow a truncated bivariate normal distribution, which in the absence of truncation would be

$$\begin{bmatrix} s_{1,k} \\ s_{2,k} \end{bmatrix} \sim \mathrm{N}(0, \mathbf{G}_0),$$

where

$$\mathbf{G}_0 = \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{1,2} & g_{2,2} \end{bmatrix}.$$

Here, $\mathbf{G}_0$ is the 2×2 (co)variance matrix between the random sire amplitude and frequency effects. The truncation points for the normal distribution are described later. Let $\mathbf{s} = [\mathbf{s}_1', \mathbf{s}_2']$ where $\mathbf{s}_i = [s_{i,1}, s_{i,2}, \ldots, s_{i,q}]'$ ($i = 1, 2$). Thus, the vector $\mathbf{s} \sim \mathrm{N}(0, \mathbf{G}_0 \otimes \mathbf{A})$, where $\mathbf{A}$ is the known additive genetic relationship matrix between sires, and q is the total number of sires in the pedigree.

As mentioned, the frequency $f_k = \omega + s_{2,k}$ of the curve must be such that

$$\frac{1}{T} \leq \omega + s_{2,k} \leq \frac{1}{2},$$

where $\frac{1}{T}$ is the bound pertaining to the slowest curve and $\frac{1}{2}$ is the upper bound for the fastest curve. These conditions had to be met during the MCMC sampling processes.

Independent scaled inverse Chi-square prior distributions were assigned to both the herd and permanent environmental variances, that is

$$\sigma_h^2 \sim \nu_h \tau_h \chi^{-2}(\nu_h, \tau_h),$$

and

$$\sigma_p^2 \sim \nu_p \tau_p \chi^{-2}(\nu_p, \tau_p),$$

where $\nu_h = \nu_p = 3$ are the degrees of freedom, and $\tau_h = 0.06$ and $\tau_p = 0.1$ are the scale parameters. The (co)variance matrix $\mathbf{G}_0$ was assumed to follow, *a priori*, the inverse Wishart distribution

$$\mathbf{G}_0 \sim \mathrm{Wishart}^{-1}(\nu_{\mathbf{G}_0}, \mathbf{V}_{\mathbf{G}_0}),$$

where $\nu_{G_0} = 4$ is the degrees of freedom and $\mathbf{V}_{G_0}$ is the scale matrix with diagonal elements equal to 0.0002 and off-diagonal element equal to 0.00001.

The joint prior density of all uncertain quantities, after augmentation with the liabilities (represented by the vector $\mathbf{l}$), was assumed to have the form

$$p(\boldsymbol{\beta}, \mathbf{h}, \mathbf{p}, \mathbf{b}, \mathbf{s}, \sigma_h^2, \sigma_p^2, \mathbf{G}_0, \mathbf{l})$$
$$= p(\boldsymbol{\beta})p(\mathbf{h}|\sigma_h^2)p(\mathbf{p}|\sigma_p^2)p(\mathbf{b})p(\mathbf{s}|\mathbf{G}_0)p(\sigma_h^2)p(\sigma_p^2)p(\mathbf{G}_0)p(\mathbf{l}|\boldsymbol{\beta}, \mathbf{h}, \mathbf{p}, \mathbf{b}, \mathbf{s}),$$

where the dependence on hyper-parameters is suppressed in the notation. The joint posterior density above can be written explicitly as,

$$p(\mathbf{l}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{p}, \mathbf{b}, \mathbf{s}, \sigma_p^2, \sigma_h^2, \mathbf{G}_0|\mathbf{y}, \theta, \omega)$$

$$\propto \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left[l_{hijkt} - \mathbf{x}'_{hijkt}\boldsymbol{\beta} - \mathbf{z}'_{h,hijkt}\mathbf{h} - \mathbf{z}'_{p,hijkt}\mathbf{p} - a_{j,k}\cos\{2\pi f_k(t-\theta)\}\right]^2\right]$$

$$\times |\mathbf{G}_0 \otimes \mathbf{A}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\,\mathbf{s}'(\mathbf{G}_0 \otimes \mathbf{A})^{-1}\mathbf{s}\right\}$$

$$\times (\sigma_h^2)^{-\frac{N_h}{2}}\exp\left(-\frac{1}{2\sigma_h^2}\,\mathbf{h}'\mathbf{h}\right)$$

$$\times (\sigma_p^2)^{-\frac{N}{2}}\exp\left(-\frac{1}{2\sigma_p^2}\,\mathbf{p}'\mathbf{p}\right)$$

$$\times (\sigma_h^2)^{-(\frac{\nu_h}{2}+1)}\exp\left(-\frac{\nu_h\tau_h}{2\sigma_h^2}\right)$$

$$\times (\sigma_p^2)^{-(\frac{\nu_p}{2}+1)}\exp\left(-\frac{\nu_p\tau_p}{2\sigma_p^2}\right)$$

$$\times |\mathbf{G}_0|^{\frac{-(\nu_{G_0}+3)}{2}}\exp\left\{-\frac{1}{2}\,\mathrm{tr}(\mathbf{V}_{G_0}\mathbf{G}_0^{-1})\right\},$$

where N is the number of cows with records, T_i is the number of periods for cow i, and N_h is the number of herds. This density is defined within the boundaries defined by the truncation points of the priors and by the parameter relationships.

The fully conditional posterior distributions of all unknowns can be derived from the joint posterior density (e.g., [1, 13, 14]). Given the liabilities, all conditional posterior distributions are recognizable, except those of the frequency parameters of the cosine function, since these enter non-linearly into the model. The vector $\boldsymbol{\beta}$ has a truncated multivariate normal distribution, and the conditional posterior distributions of both $\mathbf{h}$ and $\mathbf{p}$ are multivariate normal.

Thus, each element of $\boldsymbol{\beta}$ has a truncated univariate normal conditional posterior distribution, and each element of $\mathbf{h}$ and $\mathbf{p}$ follows a univariate normal distribution.

The "fixed" (b_j) and random parts ($s_{1,k}$) of the amplitude enter linearly in the model for liabilities, so their conditional posterior distributions can be derived from the joint posterior distribution directly. Let $\mathbf{l}^*$ be a vector of offsets with elements equal to

$$l^*_{hijkt} = l_{hijkt} - x'_{hijkt}\boldsymbol{\beta} - z'_{h,hijkt}\mathbf{h} - z'_{p,hijkt}\mathbf{p} - s_{1,k}\cos\{2\pi f_k(t-\theta)\},$$

where l_{hijkt} is liability, as before. Note that $l^*_{hijkt} = b_j\cos\{2\pi f_k(t-\theta)\} + e_{hijkt}$. Also, let $\mathbf{l}^*_j = \{l^*_{hijkt}\}$ be a vector of all offsets within the jth age-season of first calving, and $\mathbf{c}_j$ be a column vector, with elements equal to $\cos\{2\pi f_k(t-\theta)\}$ for all records in the jth age-season class. The conditional posterior distribution of the "fixed" amplitude b_j is then

$$b_j \sim \text{TN}\left(\frac{\mathbf{c}'_j\mathbf{l}^*_j}{\mathbf{c}'_j\mathbf{c}_j}, \frac{1}{\mathbf{c}'_j\mathbf{c}_j}\right).$$

The conditional posterior distribution of the random sire amplitude $s_{1,k}$ can be derived in a similar fashion. Let now $\mathbf{l}^*$ be a vector of offsets with elements equal to

$$l^*_{hijkt} = l_{hijkt} - x'_{hijkt}\boldsymbol{\beta} - z'_{h,hijkt}\mathbf{h} - z'_{p,hijkt}\mathbf{p} - b_j\cos\{2\pi f_k(t-\theta)\}.$$

Also, let $\mathbf{l}^*_k = \{l^*_{hijkt}\}$ be a vector of all offsets of daughters of sire k. In addition, let $\mathbf{c}_k$ be a column vector, with elements equal to $\cos\{2\pi f_k(t-\theta)\}$ for all records of sire k. The conditional posterior distribution of the amplitude of sire k is,

$$s_{1,k} \sim \text{N}(\hat{s}_{1,k}, (\mathbf{c}'_k\mathbf{c}_k + g^{11}A^{k,k})^{-1}),$$

where

$$\hat{s}_{1,k} = (\mathbf{c}'_k\mathbf{c}_k + g^{11}A^{k,k})^{-1}\{\mathbf{c}'_k\mathbf{l}^*_k - A^{k,k}g^{12}s_{2,k} - \mathbf{A}^{k,-k}(g^{11}\mathbf{s}_{1,-k} + g^{12}\mathbf{s}_{2,-k})\}.$$

Here, g^{ij} is the ith row and j^{th} column of matrix $\mathbf{G}_0^{-1}$; $A^{k,k}$ is the kth diagonal element of the matrix $\mathbf{A}^{-1}$; $\mathbf{A}^{k,-k}$ is the kth row of the matrix $\mathbf{A}^{-1}$ without the kth element; $\mathbf{s}_{i,-k}$, where $i = 1,2$, is the vector obtained from $\mathbf{s}_i$ by deleting its kth row.

The conditional posterior distributions of the frequency ($s_{2,k}$) of sires with progeny records do not have a closed form. A Metropolis algorithm was used

to sample these elements from their conditional posterior distributions. The algorithm is described in the following section.

The herd and cow-specific variances have scaled inverse Chi-square conditional posterior distributions. The genetic (co)variance matrix has a conditional posterior distribution that is in an inverse Wishart form. In short, Gibbs sampling can be used to draw from all recognizable conditional distributions, whereas Metropolis steps can be embedded in the scheme in order to draw the sire specific frequency effects.

2.3. Metropolis algorithm

A single-site Metropolis algorithm was tailored to sample sire-specific frequency parameters for sires with daughter records from their posterior distributions. The Metropolis implementation was done within each sire class. Univariate normal proposals were used for $s_{2,k}$, and new candidates were generated from:

$$s_{2,k}^{*} \sim N(s_{2,k}^{(p-1)}, \sigma_{s_2}^2),$$

where $s_{2,k}^{(p-1)}$ is the value of $s_{2,k}$ at the $(p-1)^{\text{th}}$ iteration of the sampler, and $\sigma_{s_2}^2$ is some fixed positive numbers leading to a reasonable rejection rate in the sampling process. Let again $l_{hijkt}^{*} = l_{hijkt} - x_{hijkt}'\boldsymbol{\beta} - z_{h,hijkt}'\mathbf{h} - z_{p,hijkt}'\mathbf{p}$, and recall that $f_k = \omega + s_{2,k}$ for all records of cows that are daughters of sire k. Then, the probability of accepting $s_{2,k}^{*}$ as a draw from its posterior can each be calculated as:

$$\alpha_{s_{2,k}} =$$

$$\min\left(1, \frac{\prod_{i=1}^{N_k}\prod_{t=1}^{T_i} \exp\left[-\frac{1}{2}[[l_{hijkt}^{*} - a_{j,k}\cos\{2\pi(\omega + s_{2,k}^{*})(t-\theta)\}]^2 + \mathbf{s}'^{*}\mathbf{G}^{-1}\mathbf{s}^{*}]\right]}{\prod_{i=1}^{N_k}\prod_{t=1}^{T_i} \exp\left[-\frac{1}{2}[[l_{hijkt}^{*} - a_{j,k}\cos\{2\pi(\omega + s_{2,k}^{(p-1)})(t-\theta)\}]^2 + \mathbf{s}'^{(p-1)}\mathbf{G}^{-1}\mathbf{s}^{(p-1)}]\right]}\right).$$

Here, N_k is the number of daughters of sire k and $\mathbf{G}^{-1} = \mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1}$.

For sires that were present in pedigree file but lacked progeny records, the fully conditional posterior distributions of their frequency parameter can be shown to be the normal processes

$$s_{2,k} \sim N((g^{22}A^{k,k})^{-1}(-g^{22}\mathbf{A}^{k,-k}\mathbf{s}_{2,-k} - g^{12}\mathbf{A}^{k,\cdot}\mathbf{s}_1), (g^{22}A^{k,k})^{-1}).$$

Above, $\mathbf{A}^{k,-k}$ is the kth row of the $\mathbf{A}^{-1}$ matrix obtaining after deleting the $A^{k,k}$ element, and $\mathbf{A}^{k,\cdot}$ is the kth row of the $\mathbf{A}^{-1}$ matrix. In addition, $\mathbf{s}_{i,-k}$ denotes a vector obtained by deleting the kth row of $\mathbf{s}_i$.

2.4. Application to clinical mastitis

2.4.1. Data

The data set represents a cohort of 36 178 cows of the Norwegian Red (NRF) breed from 5286 herds, which are the progeny of 245 sires. These cows were followed longitudinally for presence or absence of clinical mastitis until culled, or until completion of their fifth lactation, whichever occurred first. The interval ranging from 30 days prior to calving to 300 days *post-partum* was divided into five biologically meaningful periods: $(-30,0)$, $(1,30)$, $(31,120)$, $(121,210)$ and $(211,300)$ for each cow in each of the first five lactations. These periods correspond to distinct physiological stages of lactation: pre-calving, calving and start of milking, early lactation, mid-lactation and late lactation. Although the periods span different lengths of time, we interpret them as units of "biological time". The idea here is that periodicity reflects "biological" and not chronological time. The response variables were presence or absence of clinical mastitis in each of the potential 25 periods. If a cow had at least one mastitis treatment within a given period (which is rare), the record was scored as "presence", and as "absence" otherwise. Hence, each cow had at most five records on presence or absence of clinical mastitis in each lactation; a missing record occurred when a cow was culled before 300 days in the corresponding lactation. Culled cows had mastitis records for the period in which they were culled even if this period was not complete. This overstates somewhat resistance to mastitis in that period, because the opportunity of infection is smaller.

As mentioned, a cow had at most 25 binary records over the five lactations. About 72% and 45% of the cows had complete first and second lactation records, respectively. Out of the original 36 178 cows, 11 219, 5797 and 2643 cows stayed in the herd until the last period of the third, fourth and fifth lactations, respectively. Table I gives the number of cows and the corresponding mastitis incidence rates, by period, for each of the five lactations. The highest incidence rates of clinical mastitis were found for the first month after calving (periods 2, 7, 12, 17 and 22); here, incidence rate ranged from 10.1% to 18.6%. Subsequently, there was a decrease in incidence. This indicates a periodical pattern in the process of infection.

Mastitis incidences for sires with the highest and lowest amplitude and frequency along the first five lactations are presented in Figure 2. The raw series were not stationary, since incidence of clinical mastitis increased with parity. A similar pattern was observed across sires in the four groups. The first peaks were observed in the second period for almost all sires; mastitis increases after calving and then decreases, with this pattern repeated lactation after lactation.

Table I. Total number of cows and mastitis incidence, by period, for each of five lactations.

Period	Lactation	Days from calving	Number of cows	Mastitis incidence
1	1	-30–0	36 178	4.7%
2	1	1–30	36 178	10.1%
3	1	31–120	34 868	5.6%
4	1	121–210	32 644	4.4%
5	1	211–300	29 854	3.9%
6	2	-30–0	24 631	3.9%
7	2	1–30	24 631	11.5%
8	2	31–120	23 995	10.6%
9	2	121–210	22 131	7.3%
10	2	211–300	19 431	5.1%
11	3	-30–0	15 296	4.4%
12	3	1–30	15 296	15.3%
13	3	31–120	14 675	13.1%
14	3	121–210	13 123	8.3%
15	3	211–300	11 219	6.0%
16	4	-30–0	8396	4.9%
17	4	1–30	8396	17.6%
18	4	31–120	7859	14.9%
19	4	121–210	6800	9.3%
20	4	211–300	5797	6.8%
21	5	-30–0	4044	4.6%
22	5	1–30	4044	18.6%
23	5	31–120	3719	14.1%
24	5	121–210	3153	8.8%
25	5	211–300	2643	6.2%

Variation in amplitude was clearer than variation in frequency. Taller peaks were observed in the high amplitude groups of sires. The phase parameter, θ, was fixed at 2 for all series in this study, given that the first peak occurred at the second period across all series, without any apparent variation among sires. A common frequency of $\omega = \frac{1}{5}$ across all series seems evident from the figure, since the pattern repeated itself every five periods in all series.

For the model in Equation (2), the location vector β included 12 age-season of first calving classes, 5 levels of parity and 9 years of calving effects;
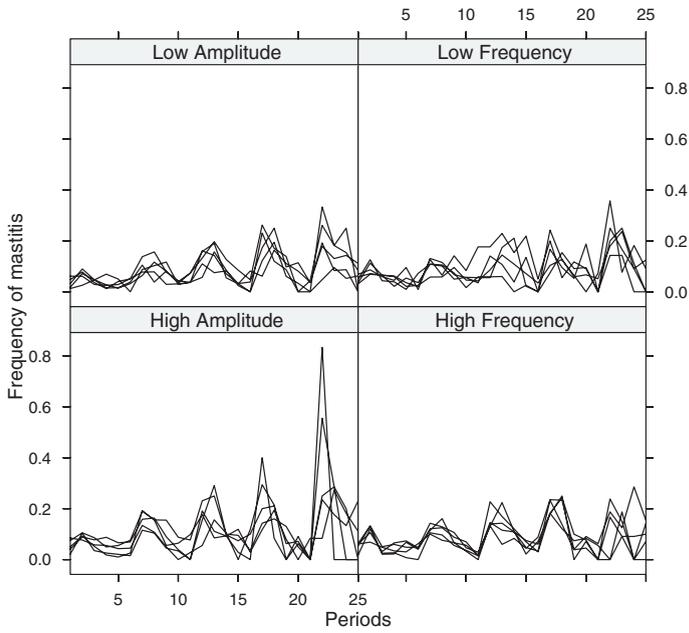
Figure 2. Mastitis incidence rate for sires with the highest and lowest amplitude and frequency along five lactations.

5286 levels of herd effects were included in **h**; the vector **p** included 36 178 cow-specific permanent environmental effects. Since there were 437 sires presented in the pedigree file, there were 437 sire-specific effects for amplitude (s_1) and for frequency (s_2).

2.5. Convergence diagnostics

A single chain of 200 000 iterations was generated. Convergence diagnostics were based on trace plots of all (co)variance components and on Raftery and Lewis [11]. Based on the diagnostics employed, 20 000 iterations were discarded as burn-in, and inferences were based on the remaining 180 000 samples. The acceptance rate of the Metropolis algorithm for sire frequency was 29.4%.

3. RESULTS

Posterior means of the "fixed" amplitudes (b_j) ranged from 0.1764 to 0.2488 over age-seasons at first calving. Larger amplitude was observed for cows that

Table II. Posterior means, standard deviations (S.D.), Monte Carlo standard errors (S.E.) and 95% posterior intervals for sire (co)variances, herd and cow variances.

Parameter	Posterior mean	Posterior S.D.	Monte Carlo S.E.	2.5%	97.5%
g_{11}	0.0068	0.0013	2.60e–5	0.0045	0.0096
g_{22}	2.34e–5	3.66e–6	4.47e–8	1.71e–5	3.14e–5
g_{12}	−0.0003	6.02e–5	1.03e–6	−0.0004	−0.0002
$r_{g_{12}}$	−0.7566	0.069	1.21e–3	−0.8666	−0.5973
σ_h^2	0.072	0.0029	3.15e–5	0.066	0.078
σ_p^2	0.086	0.0038	8.87e–5	0.079	0.094

g_{11} = variance between sires for amplitude; g_{22} = variance between sires for frequency; g_{12} = covariance between sire effects on amplitude and frequency; $r_{g_{12}}$ is the genetic correlation between amplitude and frequency; σ_h^2 and σ_p^2 are the variances between herds and between cows, respectively.

were older at first calving ($\geq$ 28 months). The posterior means, standard deviations, Monte Carlo error and 95% posterior intervals for the sire (co)variances of amplitude and frequency, and for herd and cow-specific variances are given in Table II. The posterior distributions of sire variances for amplitude and frequency are in Figure 3, and the distributions were symmetric with posterior means of 0.0068 and 0.000023 for amplitude and frequency, respectively. The posterior coefficient of variation was 15.6% for frequency and 19% for amplitude. A strong negative genetic correlation (posterior mean was −0.75) between amplitude and frequency was observed, and its posterior distribution was unimodal but slightly skewed to the left (Fig. 3(c)). Figure 3(d) shows the distribution of estimated sire amplitudes which ranged between −0.07 and 0.23. The posterior means of herd and cow-specific variances were 0.072 and 0.086, respectively. Between sire variation for amplitude and frequency was much less important than between-herd and between-cow variations.

4. DISCUSSION

If a sire has a larger value of amplitude, his daughters would be expected to have a higher risk of mastitis, due to genetic transmission of susceptibility. The estimates of frequency provide information on how frequent the curve repeats itself, and daughters of sires with higher frequency would be expected to contract mastitis more frequently. Combinations of amplitude and frequency can give information on the variation of liability to clinical mastitis over the course of several lactations. On this basis, one should select sires that,
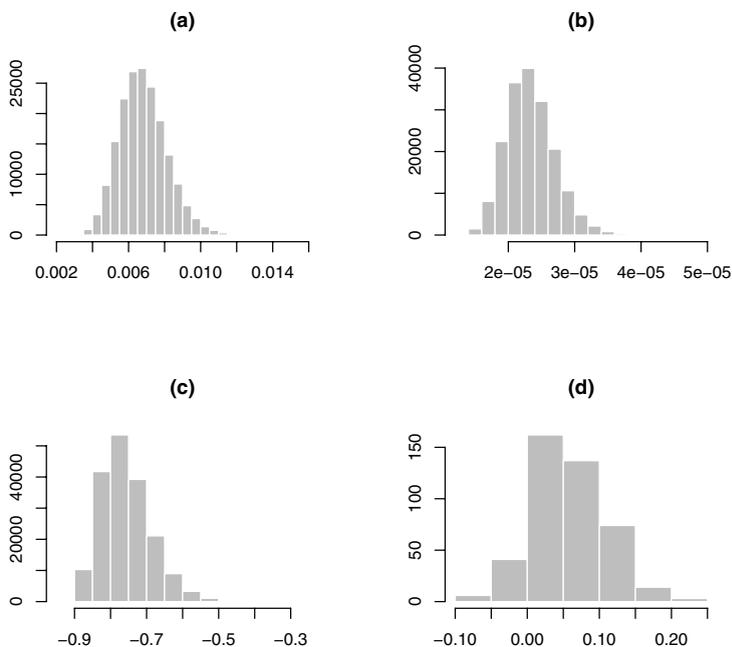
Figure 3. Posterior distributions of (a) sire variance for amplitude, (b) sire variance for frequency, (c) genetic correlation between amplitude and frequency, and (d) distribution of estimated sire amplitudes.

genetically, have smaller amplitude and frequency. However, the negative genetic correlation between amplitude and frequency suggests that if sires with smaller amplitude were selected as parents, there would be an undesirable correlated response in frequency, which would be expected to increase. Although there is small variation between sire amplitudes, the difference between sire amplitudes (ranged between −0.07 and 0.20 for sires with daughter records) indicates possible selection against susceptibility to mastitis. Most sires with negative effect on amplitude were also on the top 20 sires for resistance to mastitis when multivariate threshold models were applied to first-lactation records only [3].

The small Monte Carlo errors for herd and cow-specific variances indicate that these posterior means were estimated precisely. Our estimates agree with an earlier analysis of clinical mastitis in multiparous cows using a multivariate probit model [7].

Given the strong triangular wave pattern of mastitis incidence across lactations in this study, only a single dimension cosine function was included in the analysis. The "fixed" common frequency was set equal to a constant across all

age-season of first calving classes, based on the data. However, this assumption can be relaxed if heterogeneity in frequency is expected. For example, one would expect the series to reveal more variation if smaller intervals (larger number of periods) were defined. Higher harmonics for the cosine functions would be more likely to capture variation between animals when finer intervals are used to construct the longitudinal binary responses. Although finer intervals would reveal more variation between animals, it would also require much more intense computing time since the number of records for each animal would be larger. Further, it would also result in smaller amplitudes and weaker signals.

In addition, the phase of the series was fixed at 2, given the evidence that the first peak occurred at the second period across all series. However, it is possible to include both "fixed" and "random" phases in the model, if preferred. However, posterior intercorrelation between frequency and phase parameters inside the cosine function slows down convergence considerably. This requires further research and algorithm development.

Heringstad *et al.* [8] studied the number of cases of mastitis in first-lactation Norwegian Red cows using a censored ordinal threshold model. Repeatability ordinal threshold models could be an alternative to study multiparous mastitis records. However, the information on time to contracting mastitis would be lost. Similarly, Rodrigues-Motta *et al.* [12] studied the number of cases of mastitis in first-lactation Norwegian Red cows using Poisson and zero-inflated Poisson models. Repeatability or multivariate Poisson models would be useful to study the genetic variation of the mastitis cases in later lactations. However, more research is needed to account for censored records of cows culled early in lactation.

In conclusion, this is a first attempt at modelling longitudinal clinical mastitis using a periodic analysis. Our results suggest potential usefulness of a periodic analysis of longitudinal binary responses using sinusoidal functions if a periodically repeated pattern is present. However, we did not find evidence that genetic variation (as measured by variance between sires of cows) for amplitude or frequency is sizable in clinical mastitis.

by the Norwegian Cattle Health Service in agreement number 011.2000 by 12.10.2000. GENO Breeding and A. I. Association is acknowledged for providing pedigree information on sires.

REFERENCES

[1] Albert J.H., Chib S., Bayesian analysis of binary data polychotomous response data, J. Am. Stat. Assoc. 88 (1993) 669–679.

[2] Bolch B.W., Huang C.J., Multivariate statistical methods for business and economics, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.

[3] Chang Y.M., Gianola D., Heringstad B., Klemetsdal G., Correlations between clinical mastitis in different periods of first-lactation Norwegian Cattle using a multivariate threshold model, in: Gatsonis C., Carriquiry A., Gelman A., Higdon D., Kass R., Pauler D., Verdinelli I. (Eds.), Case Studies in Bayesian Statistics, Vol. 6, Springer-Verlag, New York, 2002, pp. 177–192.

[4] Dempster E.R., Lerner I.M., Heritability of threshold characters, Genetics 35 (1950) 212–236.

[5] Gianola D., Theory and analysis of threshold characters, J. Anim. Sci. 54 (1982) 1079–1096.

[6] Heringstad B., Klemetsdal G., Ruane J., Selection for mastitis resistance in dairy cattle – a review with focus on the situation in the Nordic countries, Livest. Prod. Sci. 64 (2000) 95–106.

[7] Heringstad B., Chang Y.M., Gianola D., Klemetsdal G., Multivariate threshold model analysis of clinical mastitis in multiparous Norwegian Dairy Cattle, J. Dairy Sci. 87 (2004) 3038–3046.

[8] Heringstad B., Chang Y.M., Andersen-Ranberg I.M., Gianola D., Genetic analysis of mastitis and number of services to conception in Norwegian Red using a censored threshold model, J. Diary Sci. 89 (2006) 4042–4048.

[9] Lippi M., Reichlin L., Common and uncommon trends and cycles, Eur. Econ. Rev. 38 (1994) 624–635.

[10] MacNeill I.B., A test of whether several time series share common periodicities, Biometrika 54 (1977) 495–508.

[11] Raftery A.E., Lewis S., How many iterations in the Gibbs sampler? Technical Report, Dept. Stat., Univ. of Washington, 1992.

[12] Rodrigues Motta M., Gianola D., Chang Y.M., Heringstad B., A zero-inflated Poisson model for genetic analysis of number of mastitis cases in Norwegian Red cows, in Proceeding of the 8th World Congress on Genetics Applied to Livestock Production, 13–18 August 2006, Belo Horizonte, Brazil, CD-ROM comm. no. 26–05.

[13] Sorensen D., Andersen S., Gianola D., Korsgaard I., Bayesian inference in threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229–249.

[14] Sorensen D., Gianola D., Likelihood, Bayesian and MCMC methods in quantitative genetics, Springer-Verlag, New York, 2002.

[15] Stoffer D.S., Detecting common signals in multiple time series using the spectral envelope, J. Amer. Stat. Assoc. 94 (1999) 1341–1356.

[16] Vahid F., Engle R.F., Common trends and common cycles, J. Appl. Econom. 8 (1993) 341–360.

[17] Wright S., An analysis of variability in number of digits in an inbred strain of guinea pigs, Genetics 19 (1934) 506.