Sequential predictions of menstrual cycle lengths

PAOLA BORTOT
Department of Statistical Sciences, University of Bologna,
via Zamboni, 33 - 40126 Bologna, Italy

GUIDO MASAROTTO, BRUNO SCARPA*
Department of Statistical Sciences, University of Padova,
via Cesare Battisti, 241 - 35121 Padova, Italy
scarpa@stat.unipd.it

SUMMARY
Forecasting the length of the menstrual cycle and of its phases is an important problem in infertility management and natural family planning. Using repeated measurements of the length of the entire cycle and of the preovular phase provided by a large English database, we describe a Bayesian hierarchical dynamic approach to the problem. A state-space process is used to model the temporal behavior of the series of lengths for each woman. The individual processes are then embedded into a multivariate system through a Bayesian hierarchy in which model parameters are allowed to vary across subjects according to a specified probability distribution. The most interesting features of the suggested method are (a) it takes into account explicitly the temporal nature of the available data and (b) if combined with a fecundability model, it can be used to forecast the probability of conception in future cycles as a function of any intercourse behavior.

Keywords: Bayesian hierarchical dynamic model; Follicular phase; Gibbs Sampler; Menstrual cycle length; State-space model.

1. INTRODUCTION
Prediction of the length of the menstrual cycle and of its phases—the preovular or follicular phase, the ovulation, and the postovular or luteal phase—is a fundamental step in infertility management (Stanford and others, 2003) and natural family planning (see, e.g. Knight and Clubb, 1996). Identification of a woman’s fertile window is widely based on calendar calculations (Ogino, 1930; Colombo and Scarpa, 1996; Lamprecht and Grummer-Strawn, 1996; Arevalo and others, 1999; Arevalo and others, 2002), sometimes combined with other self-monitoring symptoms, such as basal body temperature or vulvar observation of cervical mucus. In addition, the accuracy of some procedures, for example, postcoital tests, as well as the success of some therapies are dependent upon adequate timing of ovulation. Understanding the time evolution of the length of the entire cycle and of its phases is an important issue also in reproductive and general health studies, as menstrual dysfunction is seen to be related to decreased fertility.

*To whom correspondence should be addressed.
and increased future risk of various chronic diseases such as breast cancer, cardiovascular diseases, and diabetes (Yen, 1991; Harlow and Ephross, 1995; Guo and others, 2006).

Using repeated measurements of the duration of the entire cycle and of the follicular phase provided by a large English database, our aim is to develop a statistical model that enables prediction of each of these 2 menstrual cycle characteristics for a woman. Many statistical models have been proposed in the literature for cycle lengths. Harlow and Matanoski (1991) and Harlow and Zeger (1991) classify cycles into 2 groups, standard and nonstandard, defined as having lengths below or above 44 days, respectively, and examine covariate effects on the mean length of standard cycles through a linear mixed model. Lin and others (1997) extend this model to account for the heterogeneity between women, while Harlow and others (2000) evaluate the age effect on the probability of having a nonstandard cycle using a generalized estimating equation. Guo and others (2006) propose a mixture of a normal distribution and a shifted Weibull distribution to model both regular and anomalous cycle lengths.

All the previous references, and most of the available literature, focus on the marginal distribution of cycle length. To take explicitly into account the longitudinal nature of the data, we propose a dynamic approach for both the entire cycle length and the time of ovulation. We use a state-space process (West and Harrison, 1997) to describe the temporal behavior of the series of lengths for each woman. Our formulation is driven by the features of the phenomenon that are known from previous analyses or derived from biological considerations and inspection of the data. To capture the variability across women, the individual processes are embedded into a multivariate system through a Bayesian hierarchy in which model parameters are allowed to vary across subjects according to a specified probability distribution. The resulting Bayesian hierarchical dynamic model has the advantage of enabling a transfer of information across subjects, compensating for the relatively short history of information available on each woman and the estimation of population parameters that can be used for predictions on women not included in the database. In addition, though the model itself is non-Gaussian and nonlinear, conditionally on a suitably augmented parameter space, it becomes a hierarchical linear Gaussian state-space model on which inference can be made via the Gibbs procedures of Shephard (1994) and Carter and Kohn (1994). Thus, posterior and predictive calculations can be carried out easily and employed in applications. For example, we will show that the draws from the predictive distribution of the preovular phase length can be used to forecast the probability of conception in the next cycle as a function of coital behavior and that this possibility leads to interesting applications for the identification of the most fertile window of a woman.

The structure of the paper is as follows. Section 2 gives a detailed description of the data set. In Section 3, the Bayesian hierarchical dynamic model is developed. In Section 4, a Markov chain Monte Carlo (MCMC) algorithm for posterior calculations is discussed. In Section 5, the model is fitted to the English database and diagnostic checks are carried out. Section 6 shows how the fitted model can be used to forecast the probability of conception in future cycles as a function of the intercourse pattern. Section 7 contains a discussion of the results and some ideas for future work.

2. THE DATA

Data on 1798 women have been collected from clients of the Catholic Marriage Advisory Council of England and Wales, whose centers provided free counseling and educational service on fertility awareness and natural family planning. Women in the study, whose ages range between 18 and 50 years, were known to be healthy. Most women were also fertile, having already given birth to at least one child, and for those (about 20%) that had not yet conceived, there was no reason to doubt their fertility (Miolo and others, 1993). Each woman provided a sequence of at least 6 consecutive cycles, leading to a total of 36,641 cycles. The longest recorded sequence of consecutive cycles comprises 109 measurements. Some women contributed more than one sequence to the database, though we decided to consider only the first sequence...
available as we have no information on the reasons that caused this temporary dropout and fear that the inclusion of the following sequences might bias the analysis.

For each woman, the length of every menstrual cycle was recorded, together with the daily basal body temperature and the days during which menstrual bleeding occurred (Marshall, 1979). Following the World Health Organization standard, a menstrual cycle is defined as the interval from the first day of one bleeding episode up to and including the day before the next bleeding episode. Using the so-called “three over six rule” (Marshall, 1979), a conventional marker of the day of ovulation for each cycle was determined manually as the first time in the cycle that the minimum basal temperature over 3 consecutive days was above the maximum temperature over the 6 immediately preceding days (Miolo and others, 1993). The day of ovulation is taken as the last of the 6 days with lower temperature. However, for some cycles, the length of the preovular phase is missing either because the cycle is apparently monophasic or because the registration of the basal body temperature is incomplete. We decided to consider only sequences with at least 3 consecutive nonmissing values of the time of ovulation. Thus, the data used in the study of the follicular phase are a subset of the total data.

2.1 Data characteristics

Previous studies, biological considerations, and a rough inspection of the data set suggest that the individual sequences of cycle length and preovular phase length have some common features:

(i) As data are recorded in days, observed lengths are discrete.

(ii) A slow downward time trend is generally observed for sequences covering many years. It is well known (e.g. Treloar and others, 1967; Vollman, 1977) that the mean length tends to decrease over a long period of time, being dependent on a woman’s age. As an example, in Figure 1, the cycle length is plotted against time for the woman with the longest recorded sequence of consecutive cycles in the database.

(iii) For some long sequences, there appears to be a sudden change in the mean level, which might be due to changes in the woman’s life style or behavior.

(iv) After accounting for a possible trend, a temporal dependence among observations seems to be present for many women. For the cycle length, there appears to be a negative lag-one autocorrelation which was previously noticed also by Colombo and Bassi (1996). This suggests that for a stationary sequence, a long cycle is most likely followed by a relatively short one and vice versa. As an example, for the same woman as in (ii), Figure 1 shows the empirical autocorrelation function evaluated on the series of first differences for the cycle length.

![Fig. 1. Plot of cycle length against time and related autocorrelation of first differences for the woman with the longest sequence of consecutive cycles.](https://academic.oup.com/biostatistics/article-abstract/11/4/741/371616/1474137/616)
(v) For some women, the data include very large or very small observations that can be regarded as outliers with respect to the woman’s regular pattern. Some of these anomalous cycles might have a specific biological explanation, such as nondetectable early losses.

(vi) Heterogeneity is observed across women.

3. The model

Consider an individual woman, and let \( y_t \) denote the length in days of her \( t \)th menstrual cycle \((t = 1, 2, \ldots)\) or, alternatively, of the hypothermic phase in the basal body temperature. Our main objective is the derivation of the one-step ahead predictive distribution

\[
F_{t+1}(x) = P\{y_{t+1} \leq x | y_1, \ldots, y_t\}, \tag{3.1}
\]

for \( t = 1, 2, \ldots \). We are interested in evaluating \( F_{t+1}(x) \) in its entirety since the nonnegligible variability within individual series means that interval forecasts are more appropriate than point predictions (Marshall, 1965).

In looking for a statistical model suitable to describe the observed processes, we have to take into account that there is variability both within and between-women. The approach used to model these 2 sources of variation can be described as follows. To explain variations from cycle to cycle for a particular woman, we propose a parametric model, that is, \( P\{y_{t+1} \leq x | y_1, \ldots, y_t, \theta\} = G_{t+1}(x | y_1, \ldots, y_t, \theta) \), where \( G_{t+1} \) is fully specified and \( \theta \) is a vector of unknown parameters. We then assume that the functional form of \( G_{t+1} \) is the same for all women, while the residual variability between women can be described by allowing \( \theta \) to vary across subjects according to a probability distribution \( p(\theta | \zeta) \). Finally, we cast the problem in a Bayesian framework by specifying a prior distribution for \( \zeta \).

3.1 The individual model

We define implicitly the individual model \( G_{t+1} \) through a state-space formulation (West and Harrison, 1997). This choice is motivated by the flexibility of state-space models which, in this context, enable us to account explicitly for the temporal nature of the problem as well as for features (i)–(vi) outlined in Section 2.

Since the lengths under study measure time intervals between successive events of interest—bleeding episodes or ovulations—we can view \( y_t \) as discrete recordings of the interarrival times \( z_t \) of a continuous process, that is, \( y_t = \lfloor z_t \rfloor \), where \( \lfloor \cdot \rfloor \) indicates rounding. We then assume that the true unobserved continuous lengths \( z_t \) are generated by the model

\[
z_t = m_t + \psi_t + \lambda_t o_t, \tag{3.2}
\]

where \( m_t, \psi_t, \lambda_t, \) and \( o_t \) are defined as follows. To allow for a possible time trend in the series, we choose for \( m_t \) a random walk model:

\[
m_t = m_{t-1} + \eta_t,
\]

where \( \eta_t \) is a sequence of i.i.d. \( \mathcal{N}(0, \sigma^2_{\eta}) \) variables. After allowing for the presence of a trend, to enable autocorrelation among observations, we specify for \( \psi_t \) the ARMA(1,1) model

\[
\psi_t = \beta \psi_{t-1} + \epsilon_t + \alpha \epsilon_{t-1},
\]
where $\varepsilon_t$ is a sequence of i.i.d. $N(0, \sigma^2_\varepsilon)$ variables. Our initial choice for $\psi_t$ was an autoregressive process of order 1, but this was found to leave residual autocorrelation. Therefore, we generalized to an ARMA(1,1) model that permits more flexible forms of temporal dependence. Finally, to account for the contamination of a small fraction of outliers, we assume $\lambda_t$ to be a sequence of i.i.d. Bernoulli($\pi$) variables and $o_t$ to be a sequence of i.i.d. $N(0, \sigma^2_o)$ variables. Under this model, for a given time $t$ the variable $z_t$ has probability $\pi$ of being an additive outlier with magnitude $o_t$.

Following the state-space representation of an ARMA process given by Hamilton (1994, p. 387), model (3.2) can be written in state-space form as

$$z_t = m_t + \zeta_t + \alpha x_t + r_t,$$

(3.3)

$$m_t = m_{t-1} + \eta_t,$$

(3.4)

$$\zeta_t = \beta \zeta_{t-1} + \varepsilon_t,$$

(3.5)

$$x_t = \zeta_{t-1}$$

(3.6)

with $r_t \sim (1 - \lambda_t)\delta_0 + \lambda_t N(0, \sigma^2_o)$ i.i.d., where $\delta_0$ denotes a random variable degenerate on 0, $\lambda_t \sim$ Bernoulli($\pi$) i.i.d., $\eta_t \sim N(0, \sigma^2_\eta)$ i.i.d., and $\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$ i.i.d.. The complete individual model is then given by (3.3–3.6) together with

$$y_t = [z_t].$$

It is easy to verify that process (3.3–3.6) is parametrically identifiable and satisfies the observability condition of West and Harrison (1997, p. 144), which, broadly speaking, ensures that a state vector of dimension $k$ can be completely identified with $k$ distinct values of the mean response.

For the distribution of $z_t$, some authors propose the use of right-skewed distributions (see, e.g. Wilcox and others, 2000; Guo and others, 2006). A right-skewed distribution could be easily accommodated within the present approach, for example, with a lognormal distribution, which amounts to replacing $z_t$ in (3.3) with $\log z_t$. However, the inspection of the data, the need to capture both extremely large and extremely small observations and the diagnostic results shown in Section 5 suggest that the normal distribution is a suitable choice for $z_t$, conditional on the state process.

The random walk component (3.4) gives a stochastic description of the time evolution and provides the model with the necessary flexibility to capture both possible slow trends and the sudden changes mentioned in point (iii) of Section 2. To verify this, we tried replacing the random walk assumption with a linear trend model. The resulting model was fitted to the data allowing random effects in the intercept and in the angular coefficient to account for between-women variability, but it was found to have inferior predictive performance.

### 3.2 The population model

Most women have relatively short sequences of observations, precluding the possibility of individual estimation of the vector of model parameters $\theta = (\sigma^2_\eta, \sigma^2_\varepsilon, \pi, \sigma^2_o, \alpha, \beta)$. Our solution is a hierarchical model that allows a transfer of information across women.

The formulation of the hierarchical model follows directly from the individual model by allowing $\theta$ to vary across women according to a probability distribution $p(\theta | z)$. This way of building a hierarchical dynamic model differs from the one proposed by Gamerman and Migon (1993) and Landim and Gamerman (2000) since they construct the hierarchy on the states of the process, whereas we link the individual models through the parameter vector $\theta$. We believe that assuming a common underlying state evolution for all subjects is inappropriate in this context as it imposes unrealistic similarities between individual trajectories.
In principle, random effects could be added to each of the components of $\theta$, but in practice some simplification is necessary. In particular, the lack of information on the outlier parameters $\pi$ and $\sigma_o^2$ and the complex dependence among the ARMA parameters $\sigma_\varepsilon^2$, $\alpha$, and $\beta$ led us to assume $\pi$, $\sigma_o^2$, and $\sigma_\varepsilon^2$ as fixed across women. For $\alpha$, $\beta$, and $\sigma_\eta^2$, our most general model assumes the following mutually independent distributions

$$
\begin{align*}
\alpha & \sim N(\mu_\alpha, \sigma_\alpha^2), \\
\beta & \sim N(\mu_\beta, \sigma_\beta^2), \\
\sigma_\eta^{-2} & \sim \text{Gamma}(a_\eta, b_\eta).
\end{align*}
$$

(3.7)

In the application, however, we assess the possibility to simplify this general model by keeping one or more of the above parameters as constant across women.

In order to complete the Bayesian formulation, we need a prior specification for all model parameters. We assume a priori that parameters are mutually independent with the following diffuse distributions:

$$
\begin{align*}
\sigma_\varepsilon^{-2} & \overset{d}{=} \sigma_o^{-2} \overset{d}{=} \sigma_\alpha^{-2} \overset{d}{=} \sigma_\beta^{-2} \overset{d}{=} \sigma_\eta^{-2} \sim \text{Gamma}(0.1^3, 0.1^3), \\
\mu_\alpha & \overset{d}{=} \mu_\beta \overset{d}{=} \log(a_\eta) \sim N(0, 100), \\
\pi & \sim \text{Beta}(1, 1).
\end{align*}
$$

Checks were carried out to make sure that results were not sensitive to specific choices of prior parameters over a range that corresponds to reasonably noninformative distributions.

4. Posterior Computations

Inference is made via MCMC. Computations are greatly simplified if the parameter space is augmented to include for each woman the unobserved lengths $z_t$ and the latent Bernoulli variables $\lambda_t$ identifying outliers. Conditionally on the augmented parameter space, the model of Section 3 reduces to a hierarchical linear Gaussian state-space model for which the multi-move Gibbs Sampler algorithm of Carter and Kohn (1994) and Shephard (1994) can be used to simulate from the posterior distribution of the states. In this setting, the missing values of the follicular phase are easily handled by letting the posterior distribution of the states coincide with the priors in the recursive equations of the Kalman filter (West and Harrison, 1997). With the exception of $\alpha$, $\lambda_t$, and $z_t$, conditionally on the states, simulation of model parameters is straightforward and is carried out using standard MCMC techniques. For $\alpha$, $\lambda_t$, and $z_t$, we adopted the Gibbs sampler procedure proposed by Chen (1997). Details of the full conditional distribution and updating equation for each model parameter are given in the supplementary material available at Biostatistics online.

All programs are written in R using the sspir package (Dethlefsen and Lundbye-Christensen, 2006) for the simulation of the states. Convergence of the MCMC algorithm is checked by comparing multiple chains with dispersed starting points. However, chains show good mixing properties.

Draws from the posterior distribution are used to derive by simulation the predictive distribution $F_{t+1}$, exploiting the same augmentation device. Though we focus mainly on $F_{t+1}$, the $k$-step ahead predictive distribution can be obtained in an analogous way by forward running the Kalman filter.

5. Results and Diagnostics

The model of Section 3 is now fitted to the cycle length and the preovulatory phase length measurements from the English database. The large dimension of the database allows us to divide women randomly into 2 groups. The first group, comprising 1100 women, is used for estimation, while the second group, with 698 women, is used for diagnostics. In this way, the results obtained from the second group give an idea of the true performance of the model.

To verify whether the general population model (3.7) with random effects on each of $\alpha$, $\beta$, and $\sigma_\eta^2$ could be simplified, we also fitted models where one or more of these 3 parameters is kept constant across
women. We then compared the performance of the fitted models in terms of the mean square error (MSE) of one-step ahead predictions evaluated on the test group. Since posterior distributions for individual-specific parameters are not available for women in the test group, to derive the predictive distribution required in the computation of the MSE, these have been replaced with the relative population quantities. On the basis of this criterion, for the cycle length the optimal model was found to be the one with random effects only on $\sigma^2_\eta$ with corresponding MSE equal to 15.5, while the worst model had all parameters fixed across women and MSE equal to 16.8. For the follicular phase, the optimal model was the one with random effects on both $\sigma^2_\eta$ and $\alpha$ and MSE equal to 16.2, while the worst model had all parameters fixed and MSE equal to 17.8. For the linear trend model discussed in Section 3, the values of the MSE were 18.9 and 19.2 for the cycle length and the follicular phase length, respectively. Thus, the linear trend model has poorer predictive performance than the worst of the models based on the random walk assumption. For both menstrual characteristics, Table 1 shows posterior means and 95% credibility intervals of the population parameters of the selected models.

The estimates of $\alpha$ and $\beta$ for the cycle length model produce a negative lag-one autocorrelation which confirms the findings of Colombo and Bassi (1996), while higher-order autocorrelations are of negligible magnitude. This implies that for a stationary sequence, a long cycle is most likely followed by a relatively short one and vice versa. By contrast, estimates of $\mu_\alpha$ and $\beta$ for the follicular phase indicate the presence of a positive autocorrelation at a population level which persists at higher orders, so that consecutive preovular phases are positively dependent. The estimates of $\pi$ and $\sigma_\eta$ suggest that the inclusion of a component for capturing outlier contamination is an essential aspect of the model. For the random walk component, Figure 2 displays the histogram of posterior means of $\sigma_\eta$ evaluated for all women in the estimation group of the cycle length model. A similarly shaped histogram (not shown) is derived for the preovular phase. Figure 2 shows a spectrum of behavior across women: for the majority of women $\sigma_\eta$ is small implying near stationarity, while for a substantial minority the large values of $\sigma_\eta$ indicate strong nonstationarity.

To verify model assumptions, diagnostic checks, based on the recursive residuals defined by Frühwirth-Schnatter (1996), were carried out on the women belonging to the test group. Let $\gamma_t$, $t = 1, 2, \ldots$, be a sequence of i.i.d. Uniform $[0,1]$ random variables and $F_{t+1}(y_{t+1})$ be the predictive distribution evaluated on the actual observation $y_{t+1}$. Under a correct model specification, the recursive residuals

$$u_t = (1 - \gamma_t)F_{t+1}(y_{t+1} - 1) + \gamma_t F_{t+1}(y_{t+1}), \quad t = 1, 2, \ldots, \tag{5.1}$$

are i.i.d. Uniform on $[0,1]$. Note that definition (5.1) takes into account the discrete nature of the observed lengths $y_t$. For diagnostic purposes, it is actually easier to work with the quantile residuals (Dunn and Smyth, 1996)

$$v_t = \Phi^{-1}(u_t),$$

where $\Phi(\cdot)$ is the standard normal distribution function, so that $v_t$ is a sequence of i.i.d. $N(0,1)$ variables. Simple graphical diagnostic devices include a plot of the autocorrelation function and a normal Q–Q plot of the residuals $v_t$. For the cycle length data, these are shown in Figure 3.

Table 1. Posterior means and 95% credibility intervals for the population parameters of the selected models for the cycle length and the preovular phase length

<table>
<thead>
<tr>
<th>Length</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma_\alpha$</th>
<th>$\sigma_o$</th>
<th>$\pi$</th>
<th>$\sigma_\eta$</th>
<th>$\sigma_\eta$</th>
<th>$b_\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle</td>
<td>$-0.61$</td>
<td>$-0.012$</td>
<td>$0.96$</td>
<td>$6.86$</td>
<td>$0.076$</td>
<td>$-0.083$</td>
<td>$0.86$</td>
<td>$0.076$</td>
</tr>
<tr>
<td></td>
<td>$(-0.77, -0.45)$</td>
<td>$(-0.083, 0.058)$</td>
<td>$(0.86, 1.05)$</td>
<td>$(6.35, 7.42)$</td>
<td>$(0.064, 0.088)$</td>
<td>$-0.083$</td>
<td>$0.86$</td>
<td>$0.076$</td>
</tr>
<tr>
<td></td>
<td>$1.02$</td>
<td>$1.55$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preov</td>
<td>$-0.32$</td>
<td>$0.88$</td>
<td>$2.33$</td>
<td>$6.46$</td>
<td>$0.037$</td>
<td>$0.114$</td>
<td>$1.88$</td>
<td>$8.58$</td>
</tr>
<tr>
<td>phase</td>
<td>$(-0.37, -0.27)$</td>
<td>$(0.85, 0.91)$</td>
<td>$(2.27, 2.38)$</td>
<td>$(5.94, 7.02)$</td>
<td>$(0.031, 0.045)$</td>
<td>$0.114$</td>
<td>$1.88$</td>
<td>$8.58$</td>
</tr>
</tbody>
</table>

Downloaded from https://academic.oup.com/biostatistics/article-abstract/11/4/741/371616 by guest on 17 December 2018
Fig. 2. Histogram of posterior means of $\sigma_\eta$ for women belonging to the estimation group of the cycle length model.

Fig. 3. Autocorrelation function (left panel) and normal Q–Q plot (right panel) for the quantile residuals $v_t$ evaluated on the test data of the cycle length.

Neither the Q–Q plot nor the autocorrelation plot show serious discrepancies. This suggests that, overall, the selected model provides a satisfactory approximation of the real time evolution of the cycle length. Similar conclusions were found for the follicular phase.

As a final assessment, we verified sensitivity to cycle selection by repeating the analysis using only women who have had previous pregnancies, as they constitute a sample that should better represent the
Table 2. Sensitivity analysis: posterior means and 95% credibility intervals for the population parameters of the selected models fitted to the cycle length and the preovular phase length of women with previous pregnancies.

<table>
<thead>
<tr>
<th>Length</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma_{\alpha}$</th>
<th>$\sigma_{\beta}$</th>
<th>$\pi$</th>
<th>$\sigma_{\pi}$</th>
<th>$a_{\eta}$</th>
<th>$b_{\eta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle</td>
<td>-0.53</td>
<td>0.032</td>
<td>1.06</td>
<td>6.59</td>
<td>0.085</td>
<td>—</td>
<td>1.07</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>(-0.75, -0.31)</td>
<td>(-0.102, 0.170)</td>
<td>(0.79, 1.33)</td>
<td>(5.89, 7.29)</td>
<td>(0.069, 0.105)</td>
<td>—</td>
<td>(0.92, 1.25)</td>
<td>(1.29, 2.13)</td>
</tr>
<tr>
<td>Preovular phase</td>
<td>-0.32</td>
<td>0.89</td>
<td>2.34</td>
<td>6.47</td>
<td>0.043</td>
<td>0.144</td>
<td>1.98</td>
<td>9.43</td>
</tr>
<tr>
<td></td>
<td>(-0.39, -0.25)</td>
<td>(0.85, 0.92)</td>
<td>(2.25, 2.42)</td>
<td>(5.68, 7.28)</td>
<td>(0.033, 0.056)</td>
<td>(0.081, 0.202)</td>
<td>(1.62, 2.41)</td>
<td>(7.04, 11.97)</td>
</tr>
</tbody>
</table>

population of fertile women. Table 2 shows posterior means and 95% credibility intervals for the population parameters derived from the reduced sample. Results are in line with those of the previous analysis, though credibility intervals are wider due to the smaller sample size, which for both menstrual characteristics is around 45% of the original size.

6. USING THE FITTED MODEL

The aim of this section is to illustrate some possible applications of the developed model. Figure 4 displays the observed cycle lengths for a particular woman together with the 95% predictive intervals derived from $F_{t+1}(x)$. It should be borne in mind that, in practice, the output is displayed gradually, evolving from left to right as successive observations become available and that the prediction interval for the length of the $(t+1)$th cycle is computed at the end of the $t$th cycle.

The first panel of Figure 5 shows the observed lengths of the preovular phase for a different woman, while the bottom panel shows the predictive densities of the preovular phase length for her 6th and 31st cycle. The 2 densities are rather different. In particular, the 31st cycle density is shifted to the left and is less dispersed. The location shift shows how the model learns from the data the local level of the series, which in Figure 5 (top panel) appears to decrease. The reduction in dispersion is due to the extra data that are available by the time of the 31st cycle.

The predictive distribution of the preovular phase length can be used directly in natural family planning and infertility management. However, for this type of application, we also need the probability of conception in a cycle given a particular intercourse behavior. Let $R$ be a vector describing a particular coital pattern during the cycle, that is, $R = [r_1, r_2, \ldots]$, where $r_i = 1$ if an intercourse act occurs on the $i$th day and $r_i = 0$ otherwise, and $U(y, R)$ be the probability of conception in the cycle given the day of ovulation, $y$, and the intercourse behavior $R$. An estimate of $U(y, R)$ can be derived from Colombo and Masarotto (2000) who applied the Schwartz and others (1980) fecundability model to a large European data set to estimate day-specific probabilities of conception.

The function $U(y, R)$ cannot be used prospectively since we do not know the day of ovulation in the next cycle. However, if we combine $U(y, R)$ with the predictive distribution of $y$, we can compute the probability of conception in the next cycle, $V_{t+1}(R)$, given the intercourse behavior $R$ and the previous observed lengths of the preovular phase through

$$V_{t+1}(R) = \int U(y, R)dF_{t+1}(y). \quad (6.1)$$

Equation (6.1) has interesting applications. For example, if we want to determine the $x$ most fertile days during the next cycle, we can simply find $R$ which maximizes $V_{t+1}(R)$ under the restriction $r_1 +
Fig. 4. Observed cycle lengths with 95% prediction intervals for an individual woman.

Fig. 5. Observed lengths of the preovular phase for an individual woman (top panel) and predictive densities for her 6th and 31st cycle (bottom panel).
An example of these computations is given in Figure 6 in which we report the most fertile intervals of length $1, 2, \ldots, 12$ for the 6th and the 31st cycle of the woman considered in Figure 5. Observe that, as a consequence of the differences in the location of the predictive densities displayed in Figure 5, the most fertile intervals for the 31st cycle are shifted to the left. The identification of the most fertile days is often used by couples attempting to achieve or avoid conception through fertility awareness methods. Generally, a loss function including a penalty for large numbers of abstinence days can be defined in order to identify rules having a low or high pregnancy rate as appropriate (Scarpa and Dunson, 2007).

As another example of the use of $V_{i+1}(R)$, for the same woman considered in Figure 5, we show in Figure 7, the risk of conception when intercourse occurs only on the first $i$ ($i = 1, \ldots, 12$) days of the cycle with one of the following 2 regimes: daily intercourse and intercourse only on odd-numbered days. The daily intercourse assumption provides an upper bound for the probability of conception in a cycle and can be used to estimate the length of the infertile preovular period.

Since only point estimates of $U(y, R)$ are given in Colombo and Masarotto (2000), the results shown in Figures 6 and 7 are based on a plug-in method, where $U(y, R)$ is replaced by its estimate in (6.1). For a fully Bayesian analysis, which properly accounts for uncertainty, the posterior distribution of $U(y, R)$ would be required.

![Fig. 6. Most fertile intervals for cycles 6 and 31 of the woman considered in Figure 5. The height of the bars gives the predicted probability of conception if there are intercourse acts every day in the interval indicated at the bottom of the bars.](https://academic.oup.com/biostatistics/article-abstract/11/4/741/371616)
Fig. 7. Prediction of the probabilities of conception for cycles 6 and 31 for the woman considered in Figure 5 under the assumptions that she had intercourse every day and on odd days in the first $i (i = 1, \ldots, 12)$ days of the cycle.

7. DISCUSSION

We have proposed a model for the temporal structure of sequential observations of menstrual cycle characteristics based on a state-space formulation that lends itself naturally to prediction. Diagnostics show that the model does well overall in the complex task of describing the dynamics of the length of the entire cycle and of its preovular phase. In particular, the formal identification of a component of autocorrelation gives an important insight into the biological process connected to menstrual cycle. A side product of the analysis is the derivation of the probability of occurrence of a nonstandard cycle. This is obtained from the posterior distribution of $\lambda_t$ and is useful in reproductive health studies for detecting ovarian dysfunctions (Guo and others, 2006).

The model can be adapted to include additional information when available, such as the woman’s age at the beginning of the observation. A term depending on the covariates $X$, say $X\phi$ where $\phi$ is a vector of unknown parameters, can be added in (3.3), and $\phi$ can be estimated through a further step of the MCMC procedure. Unfortunately, in our study age is unavailable for most women, which precludes the evaluation of its contribution to the overall heterogeneity.
Another possible extension of the proposed methodology is a bivariate joint model for cycle and follicular phase lengths. As these 2 menstrual characteristics are likely to be dependent, such a model might benefit from a transfer of information between the 2 processes with resulting improved estimates. A simple approach to tackle this issue is to use a regression model of the postovular phase length on the cycle length, where the postovular phase length is obtained as the difference between the total length and the preovular phase length. We work in this context with the postovular phase instead of the preovular phase because the biomedical literature suggests that the lutheal phase is much more regular than the follicular one (Ogino, 1930; Knaus, 1934). In fact, by a probably oversimplified rule, women with regular cycles interested in avoiding pregnancy are advised to identify their abstinence period based on an assumed 14-day postovular phase. This greater regularity of the lutheal phase should simplify dependence specification. Second, if such a regression model were able to capture the relationship between the 2 menstrual features, then there would be a cut in the full likelihood between the component associated with the parameters of the marginal distribution of the cycle length and the component associated with the dependence parameters, which would imply that no gain is obtained from a joint model in terms of either estimation efficiency or prediction power. We found that satisfactory fits are obtained from the simple model \( \log(\text{postovular length}) = \alpha + \beta \log(\text{cycle length}) + \epsilon \), with random effects added to the intercept and the angular coefficient to pick up variability across women. Diagnostics on the residuals of this model reveal no serious deviations from the classical assumptions of independent and normally distributed errors, leading us to believe that a joint construction would not lead to a substantial benefit over the marginal analysis of the previous sections.

For use in natural family planning, the model estimated in Section 5 could be implemented in a simple electronic device to be offered to couples. Taking as an input the previously observed lengths, which the couple would record sequentially, the electronic tool could provide prediction of the next cycle length, and time of ovulation as well as the daily probabilities of conception derived by programming the method described in Section 6. However, such a tool could only be used by relatively healthy couples, as it would be based on estimates from a healthy cohort.

Although we limited our study to the prediction of the length of the cycle and of its follicular phase, it should be noted that the same ideas can be adapted to the forecasting of other menstrual parameters, for example, the number of days before the peak in the cervical mucus or the length of the bleeding period (Harlow and others, 2000).

**SUPPLEMENTARY MATERIAL**

Supplementary material is available at http://biostatistics.oxfordjournals.org.

**ACKNOWLEDGMENTS**

The authors would like to thank Bernardo Colombo for his helpful advice, his encouragement and for providing the data, John Marshall who started and managed the collection of the data set for many years at the Catholic Marriage Advisory Council in London and Stuart Coles and David Dunson for their suggestions on a preliminary version of the paper. We are also grateful to the referees for many insightful comments. Conflict of Interest: None declared.

**REFERENCES**


Cycle lengths predictions


[Received February 6, 2008; revised October 7, 2009; accepted for publication October 20, 2009]