Modelling Menstrual Cycle Length in Athletes Using State-space Models

Thiago Oliveira
The Insight Centre for Data Analytics, NUI Galway, Ireland

Georgie Bruinvels
Orreco, Galway, Ireland

Charles Pedlar
Orreco, Galway, Ireland

John Newell (john.newell@nuigalway.ie)
School of Mathematics, Statistics & Applied Mathematics, National University of Ireland Galway, Ireland

Research Article

Keywords: menstrual cycle length, athletes, space models, basal body temperature (BBT)

Posted Date: December 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-122553/v1

Version of Record: A version of this preprint was published at Scientific Reports on August 20th, 2021. See the published version at https://doi.org/10.1038/s41598-021-95960-1.

Modelling menstrual cycle length in athletes using state-space models

Thiago de Paula Oliveira[1,2,3,**], **Georgie Bruinvels**[2], **Charles Pedlar**[2], **and John Newell**[1,3,*]

[1]School of Mathematics, Statistics & Applied Mathematics, National University of Ireland Galway, Ireland
[2]Orreco, Galway, Ireland
[3]The Insight Centre for Data Analytics, NUI Galway, Ireland
[*]john.newell@nuigalway.ie
[**]thiago.oliveira@ed.ac.uk

ABSTRACT

The ability to predict menstrual cycle length to a high degree of precision enables female athletes to track their period and tailor their training and nutrition correspondingly knowing when to push harder when to prioritise recovery and how to minimise the impact of menstrual symptoms on performance. Such individualisation is possible if cycle length can be predicted to a high degree of accuracy. To achieve this, a hybrid predictive model was built using data on 16,990 cycles collected from a sample of 2,178women (mean age 33.89 years, range 14.98 - 47.10, number of menstrual cycles ranging from 4 - 53). To capture the within-subject temporal correlation, a mixed-effect state-space model was fitted incorporating a Bayesian approach for process forecasting to predict the duration (in days) of the next menstrual cycle. The modelling procedure was split into three steps(i) a time trend component using a random walk with an overdispersion parameter, (ii) an autocorrelation component using an autoregressive moving-average (ARMA) model, and (iii) a linear predictor to account for covariates (e.g. injury, stomach cramps, training intensity). The inclusion of an overdispersion parameter suggested that 26.81% $[24.14\%, 29.58\%]$ of cycles in the sample were overdispersed where the random walk standard deviation under a non-overdispersed cycle is 1.0530 $[1.0060, 1.0526]$ days while under an overdispersed cycle it increased to 4.7386 $[4.5379, 4.9492]$ days. To assess the performance and prediction accuracy of the model, each woman's last observation was used as test data. The Root Mean Square Error (RMSE), Concordance Correlation Coefficient (CCC) and Pearson correlation coefficient (r) between the observed and predicted values were calculated. The model had an RMSE of 1.6126 days, a precision of 0.7501 and overall accuracy of 0.9855. In the absence of hormonal measurements, knowing how aspects of physiology and psychology are changing across the menstrual cycle has the potential to help female athletes personalise their training, nutrition and recovery tailored to their cycle to sustain peak performance at the highest level and gain a competitive edge. In conclusion, the hybrid model presented here is a useful approach for predicting menstrual cycle length which in turn can be used to support female athlete wellness to optimise performance

Introduction

The availability of mobile apps developed to track the menstrual cycle is growing as they are becoming increasingly popular to use for contraception purposes, fertility awareness and exercise planning. These apps can be grouped broadly as calendar-based, basal body temperature (BBT), or symptothermal[1–3]. Calendar apps generally use simple algorithms based on empirical measurements to predict cycle phase length[4]; BBT apps describe a woman's menstrual variation through her basal body temperature rise[5] while symptothermal apps measure parameters such as cervical mucus changes, bleeding period, sexual intercourse, etc[2].

The mobile app that generated the data used in the study is called FitrWoman, a free calendar-based app that enables users to track their menstrual cycle and symptoms, alongside the provision of information about wellness, nutrition and exercise, based on the athlete's predicted menstrual cycle phases and length. The target audience is female athletes who wish to track their menstrual cycle length in order to improve their performance. As a woman's body may respond and adapt differently throughout their cycle, different planning and preparation over the menstrual cycle phases[6] are required. The user inputs daily information on 25 symptoms variables such as flow, bloating, constipation, injury, illness, irritability, weakness, etc.

As few apps are accurate in terms of menstrual cycle length prediction[7], the development of an appropriate accurate parametric model for one-step-ahead forecast cycle length is required. Such a model should take into account the between and within-woman variability to identify menstrual cycle patterns and how each symptoms could affect cycle length, alongside the implications of significant alterations in cycle length

According to several studies[8–11], the menstrual cycle length can be classified into two groups 'standard' and 'menstrual

dysfunction', where a cycle length greater than 35 days is classified as 'menstrual dysfunction' and otherwise as standard. Many statistical models have been proposed in the literature in order to describe these different groups of menstrual cycles[2,12–15]. Generally, cycle length related to the 'standard' group can be analysed using classical statistical approaches while the mixture of standard and non-standard cycles can be analysed using a mixture distribution to account for the major symmetric distribution and for the component corresponding to the heavy right tail[11,12]. To account for the within individual variability we focused on the dynamic aspect of menstrual cycles over time, as discussed by ref.[13], who derived a predictive distribution based on individual repeated measurements using a state-space model formulation. According to these authors, state-space models under a Bayesian approach have the advantage of incorporating between subject information to compensate for the relatively large number of subjects with a low quantity of repeated measurements, and to make predictions for women not included in the sample.
It is well established that having a regular menstrual cycle is a 'vital sign' demonstrating that the body is likely to be in an adaptive state and is tolerating the physical and psychological stressors that are being placed on it[16]. Significant elongations in cycle length are associated with adverse health and fertility outcomes[17–20], therefore gaining a better understanding of the interrelating risk factors for cycle length extension is important.

In this paper, the first objective was to develop an appropriate parametric state-space formulation for the marginal distribution of standard menstrual cycles for female athletes. In addition, symptom variables were included in the model's linear predictor in order to evaluate how the individual reported symptoms may affect an athlete's menstrual cycle duration. The second aim was to develop a one-step ahead forecasting interval approach, based on a state-space formulation, to describe the observed and state process while taking into account both between and within-woman variability.

Results and Discussion

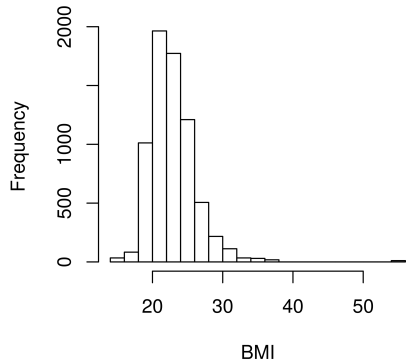
Results from the state-space models, state-space mixed-effects models and linear mixed-effects models (LMM), fitted using the available data, are summarised in Table 1. In general, the Bayesian information criteria (BIC) suggests that the random walk models fitted better than the LMM models when modelling menstrual cycle length, in agreement with the results reported by ref.[13] while contradicting the results of[2] who report an $R^2 = 0.99$ when fitting a simple linear regression.
The inclusion of r_{ij} to model overdispersed cycles lengths was fundamental to describe menstrual cycle dynamics as evidence by the BIC criteria where a reduction of $\approx 51.30\%$ compared to $y_{ij} = m_{ij} + \epsilon_{ij}$, and 54.37% compared to $y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, Age_{ij} + \epsilon_{ij}$ is evident, as shown in Figure 1. Additionally, the inclusion of a moving-average (MA) parameter was necessary to capture the dynamism of shorter cycles followed by longer cycles and vice-versa. In summary, a random walk with a random variable to capture overdispersion r_{ij} plus a MA(1) model demonstrated the best fit to the data.

Table 1. Model selection for stages I and II. Number of parameters (N. Par.); root mean square error (RMSE), concordance correlation coefficient (CCC), and Pearson correlation coefficient (r) between fitted values and test data for forecast, and Bayesian information criterion (BIC) based on the model. Selected model, lowest RMSE and BIC where higher CCC and r are in bold.

Model	N. Par.	Forecasting			BIC
		RMSE	CCC	r	
$y_{ij} = m_{ij} + \epsilon_{ij}$	3	1.6104	0.7433	0.7606	16,961.19
$y_{ij} = m_{ij} + \text{AR}(1)$	4	1.6740	0.7546	0.7618	9,733.27
$y_{ij} = m_{ij} + \text{MA}(1)$	4	1.6640	0.7564	0.7637	9,726.88
$y_{ij} = m_{ij} + \text{ARMA}(1,1)$	5	1.6782	0.7560	0.7630	9,314.73
$y_{ij} = m_{ij} + r_{ij} + \epsilon_{ij}$	5	1.6153	0.7362	0.7479	8259.67
$y_{ij} = m_{ij} + r_{ij} + \text{AR}(1)$	6	1.6098	0.7410	0.7510	7797.06
$y_{ij} = m_{ij} + r_{ij} + \text{MA}(1)$	6	1.6126	0.7401	0.7501	7739.40
$y_{ij} = m_{ij} + r_{ij} + \text{ARMA}(1,1)$	7	1.6031	0.7405	0.7521	7995.92
$y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, Age_{ij} + \epsilon_{ij}$	5	1.6481	0.7265	0.7457	18,099.64
$y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, Age_{ij} + \text{AR}(1)$	6	1.7796	0.6964	0.7087	13,314.69
$y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, Age_{ij} + \text{MA}(1)$	6	1.7608	0.7005	0.7137	13,381.90
$y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, Age_{ij} + \text{ARMA}(1,1)$	8	1.7769	0.6998	0.7113	13,289.62

To assess model performance, we compared the forecasts of these models using the RMSE of one-step ahead predictions, CCC and Pearson correlation coefficient evaluated on the test group. Table 1 demonstrates that better forecast predictions are evident using a random walk rather than a LMM model and that there is not much difference

between the random walk models in terms of forecasting. As a consequence, the BIC criteria can be used to select the error structure. After selecting the trend and error structures, the next stage of the analysis was the selection of potentially useful explanatory variables from the set of 28 potential explanatories representing a variety of reported symptoms by the i-th woman, including an interval-based variable representing a woman's body mass index (Kg/m^2) (Table 2), classified as discussed by ref.[21]. In this analysis, underweight classes I and II are classified as 'severely' and 'very severely underweight' while obese classes I, II, and III represented moderately, severely and very severely obese respectively. The sample of women had a reported BMI of between 14.44 and 54.25, with a mean of 22.85 Kg/m^2; the absolute frequency is shown in Figure 1.



	BMI (Kg/m^2)	
Category	From	To
Underweight II		15
Underweight I	> 15	16
Underweight	> 16	18.5
Normal	> 18.5	25
Overweight	> 25	30
Obese Class I	> 30	35
Obese Class II	> 35	40
Obese Class III	> 40	

Figure 1 & Table 2. Histogram with BMI absolute frequency; and classification based on body mass index (BMI)

The selected state-space model with posterior means and 95% credibility intervals for the population parameters (after predictor selection) is presented in Table 3. As the model parameterisation used facilitates the interpretation

Table 3. Posterior means and 95% Bayesian credibility interval for $\boldsymbol{\Theta}$

			95% Credible Interval	
Parameter	Estimate	SE	Lower	Upper
β_0	26.8974	0.0639	26.7704	27.0197
π	0.2681	0.0139	0.2414	0.2958
θ_0	-0.0669	0.0307	-0.1287	-0.0088
α_1 (Heavy Legs)	-0.0619	0.0363	-0.1320	0.0084
α_2 (Increased Breathing)	0.1240	0.0447	0.0353	0.2114
α_3 (Injury)	0.1438	0.0486	0.0498	0.2382
α_4 (Stomach Cramps)	0.1315	0.0303	0.0717	0.1911
α_5 (Tender Breasts)	-0.1255	0.0248	-0.1734	-0.0767
α_6 (Flow Amount: Heavy)	0.0881	0.0410	0.0061	01679
α_7 (Flow Amount: Medium)	0.1462	0.0140	0.1199	0.1742
α_8 (Flow Amount: Light)	0.0940	0.0302	0.0357	0.1544
α_9 (Flow Amount: Spotting)	0.0977	0.0330	0.0332	0.1611
α_{10} (Flow Amount: None)	0.0308	0.0131	0.0051	0.0565
σ_w	4.7386	0.1058	4.5379	4.9492
σ_η	1.0530	0.0241	1.0060	1.0526
σ_ϵ	1.5702	0.0428	1.4864	1.6545

of the role played by the explanatory variables our analysis reveals important insights on how some symptoms affect menstrual cycle length. We found that the overall menstrual cycle length without any reported symptoms was around 26.89 [26.77, 27.02] days. When considering the average number of times each symptom was reported (heavy legs= 0.18, increased breathing= 0.09, stomach cramps= 0.42, injury= 0.05, tender breast= 0.29, and flow amount none=0.49, spotting= 0.32, light= 0.55, medium= 2.95 and heavy= 0.32) the overall menstrual cycle increased from 26.89 to 27.48 days in length, on average, in agreement with ref.[12] and ref.[2].
Additionally, the reporting of increased breathing, injury, stomach cramps and flow amount was associated with increased menstrual cycle length while the reporting of heavy legs and tender breasts was associated with decreased cycle length. As example if a woman reported tender breasts 10 times over her cycle, as consequence her predicted

menstrual cycle length is estimated to reduce, on average, by $0.1255 \times 10 = 1.255$ days. Clearly, self-track symptoms depend on both user engagement, app design, and avoid ambiguous language to describe each level of a symptom (such as flow amount light and spotting). Consequently, to turn it more consistent filtering the original database based on scientific literature is essential way to reduce the effect of bias in the covariates used to fit the model, as described by ref.[11].

The estimated value of π suggests that the probability of a non-standard (overdispersed) menstrual cycle length occurring in this population of interest is 0.2681. As a consequence we can infer that 26.81% [24.14%, 29.58%] of cycles in the sample are overdispersed. Furthermore, while a non-overdispersed cycle had an standard deviation (SD) of $\sigma_\eta = 1.0530$ [1.0060, 1.0526], the SD of an overdispersed cycle increases where $\sigma_w = 4.7386$ [4.5379, 4.9492], which represents a 4-fold increment. According to ref.[22], between and within-variability in cycle characteristics should emphasising as an important health indicator to assess behavioural, metabolic, and environmental factors. Therefore, the inclusion of θ and σ_w play an essential rule in the proposed model, as can be easily seen when considering Figure 2. This Figure shows the probability that the proposed model (3) considers an observation as overdispersed where

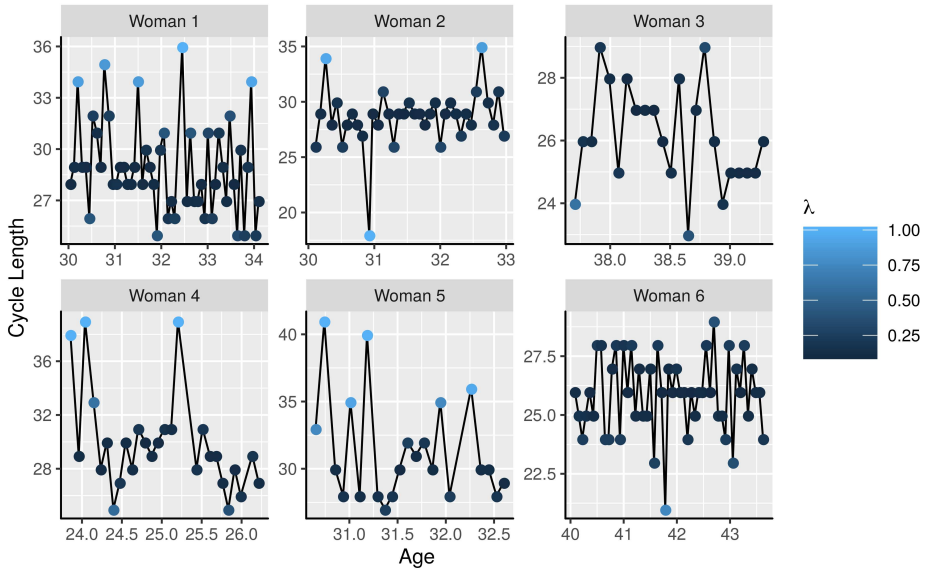


Figure 2. Example of six women profiles showing the probability that the proposed model considers an observation as overdispersed, where λ represents the probability of λ_{ij} being equal 1 for a given observation

the results clearly demonstrate that $r_{ij} = \lambda_{ij} w_{ij}$ is capturing menstrual cycles with overdispersion.

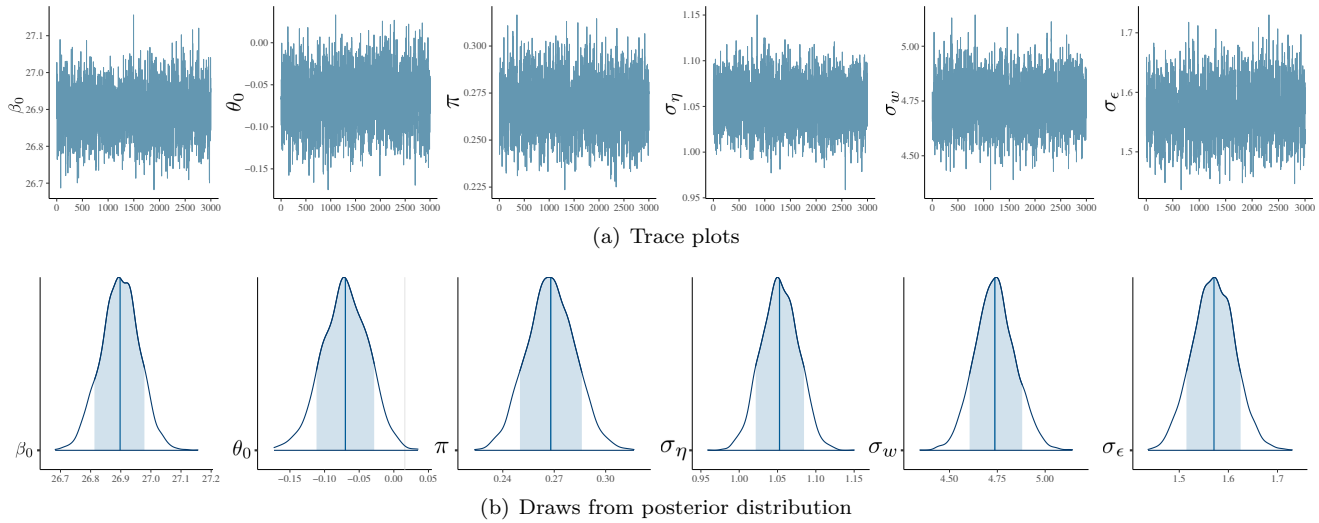
Using this model, knowledge and understanding can be gleaned as to how symptom variables affects the menstrual cycle which is clearly important for individual athletes, coaches and healthcare professionals. Furthermore, these results can improve the forecasting intervals helping women know more about their own body and cycles based on the occurrence of symptoms during a particular phase of their cycles. Although information relating to follicular and luteal phases was not available in the data, a strong linear correlation between menstrual cycle length and follicular phase was reported by ref.[23–25] where the correlation tended to increase with age. To predict ovulation time, further studies which include both luteal and follicular phases and basal body temperature (BBT) are needed to extend the proposed model[2] .

Although an ARMA(1,1) model was not needed in this analysis, we have demonstrated that some women have a positive lag-one autocorrelation while others have a negative lag-one autocorrelation. These results contradict the findings of[13,26] who report a small general negative autocorrelation for a woman's profiles. In order to better investigate the variability of an autoregressive coefficient, we modified the state-space formulation to accommodate this source of random variation by assuming that $\phi_i = \phi_0 + \phi_{0i}$, with $\phi_{0i} \sim N\left(0, \sigma_\phi^2\right)$. However, the normal assumption for ϕ_{0i} was not justified as the normal Q-Q plot suggested a distribution with heavy tails and asymmetry; as a consequence 80% of points were outside of the 95% simulated envelopes for this random effect (Figure S1).

We also observed that some women had a long cycle followed by a short one and vice versa as observed by ref.[13]. However, we found while $\hat{\theta} = -0.0669$ with $\text{CI}_{95\%} : [-0.1287, -0.0088]$ the estimate of the same parameter described by ref.[13] was -0.61 $[-0.77, -0.45]$. It appears that the sample of female athletes that these analyses are based on had

more regular menstrual cycles than a sample of 1,798 women observed from clients of the Catholic Marriage Advisory Council of England and Wales. Although we have a higher number of women in our sample than in[13], the time series in their sample were longer (up to 109 measurements) compared with up to 55 measurements in this sample. In order to account for the between-subject variability, we included a random effect in the moving-average coefficient given by $\theta_i = \theta_0 + \theta_{0i}$, with $\theta_{0i} \sim N\left(0, \sigma_\theta^2\right)$. However, we observed the same problem as reported when considering the autoregressive coefficient where more than 70% of points were outside of the 95% simulated envelopes, lower asymmetry compared with ϕ_{0i} and heavy tails (Figure S2). Therefore, to avoid the inclusion of bias in individual forecasting predictions these random effects were dropped from the model. Clearly, further work is needed in order to accommodate individual estimation for the autocorrelation and moving-average coefficients to improve model performance at the individual level.

The analysis workflow was as follows: we initially checked the Bayesian assumptions and the posteriori distribution through the use of suitable plots of the Markov chain Monte Carlo (MCMC) draws from the posterior distribution, Gelman-Rubin diagnostic and autocorrelation plots of all parameters in the model. Figure 3(a) shows the iterates of β_0, π, θ_0, σ_η, σ_w, and σ_ϵ after a burn-in of 10,000 simulated iterations, which indicates convergence of the chains and stationary distributions, as the samples appear to be randomly sampled from the same region of the y-axis rarely venturing outside that area. The autocorrelation and Gelman-Rubin statistics[27] were used to assess model convergence. The results suggest that the autocorrelation doesn't drop dramatically from lag 0 to 50 (Figure S3) suggesting a moderate to high autocorrelation among samples. In order to reduce the impact of this problem we stipulated a thinning of 50. On the other hand, the Gelman-Rubin statistic based on 3 chains shows all upper 95% confidence intervals were below 1.01 meaning the chains have converged. Figure 3(b) shows the posterior



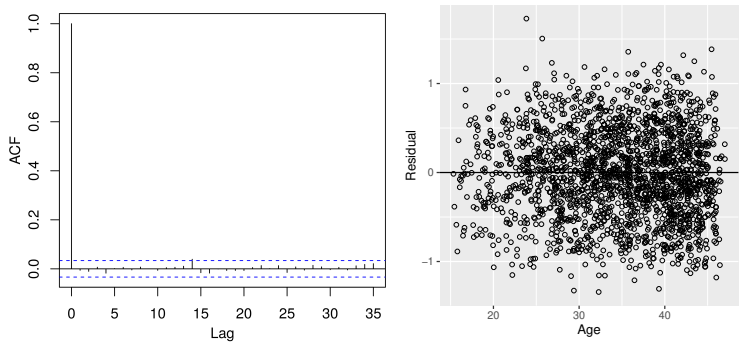
(a) Trace plots

(b) Draws from posterior distribution

Figure 3. (a) Trace plots of Markov chains and (b) Markov chain Monte Carlo (MCMC) draws from the posterior distribution of the parameters β_0, θ_0, π, σ_η, σ_w, and σ_ϵ based on a sample of length 3000

densities obtained for estimated parameters derived from 3 Markov chains with 3.000 samples per chain leading a computational time around 23 hours executed on Dell Inspiron 17 7000 with 10th Generation Intel® Core™ i7 processor, 1.80GHz × 4 processor speed, 16GB random access memory (RAM) plus 20GB of swap space, 64-bit integers, and the platform used is a Linux Mint 19.2 Cinnamon system version 5.2.2-050202-generic. In summary, the posterior distribution has been well characterised by the drawn samples as no unexpected peaks or strange shapes in the posterior density were observed that could be a sign of poor model convergence. As a final assessment, the autocorrelation function was checked (Figure 4), which shows no serious discrepancies or patterns that warrant attention.

Once the assumptions were verified, we then start to evaluate the agreement between fitted values and observed values as well as forecast intervals. Figure 5 shows the fitted curves for menstrual cycle length of six women, their 95% credible interval, and the one-step-ahead point forecast with 80%, 95% and 99% forecast intervals. We observed that the random walk with overdispersion parameter and MA(1) model does well in describing the complex dynamics menstrual cycle length over time, and this conclusion is underpinned by the residual diagnostic, and high values of CCC and Pearson correlation between fitted and observed values by woman. This results also shows that linear



(a) ACF (b) Residual vs. Age

Figure 4. (a) Residual autocorrelation plot, and (b) residual versus age

or linear mixed-effects models should not be applied to explain the variability of menstrual cycle length, as they generally do not follow the necessary assumptions of linearity. However, a study done in 2019 by ref.[2] is still using linear models in order to explain cycle length observed from a large database of cycles collected through an app. This authors shows an $R^2 = 0.987$ without any discussion about if their model assumptions are fulfilled, nor a residual diagnostics, and, as explained by ref.[28], a high R^2 values do not necessarily means that regression model provides an adequate fit to the data or imply that the regression model will be an accurate predictor.

In order to show the improvement in point estimate, credible interval, and forecasting interval when we include $r_{ij} = \lambda_{ij} w_{ij}$ into the model to describe cycle length, we built the Figure 6 using basically the same model, but dropping r_{ij} out the model. Results show that when we add r_{ij} in the model, is expected an increment in the values of Pearson correlation and concordance correlation coefficient mainly for women who have more overdispersed cycles, better forecast prediction, and a shrinkage in the credible interval.

Finally, to evaluate the one-step-ahead point forecast prediction we introduce a new group, as test data, starting with 1,029 women whom have at least 3 repeated measurements and the results is shown in the Figure 4. Due to

Table 4. Evaluation of one-step-ahead forecast prediction based on root mean square error (RMSE), concordance correlation (CCC), Pearson correlation (r), and accuracy (C_b) coefficients between the predict and observed values of a new group with N women whom have n_i cycles

			CCC				
N	n_i	RMSE	Est	Lower	Upper	r	C_b
1029	3	5.2654	0.2544	0.2072	0.3005	0.3066	0.8297
760	4	5.0889	0.3113	0.2567	0.3640	0.3657	0.8514
603	5	5.1767	0.2670	0.1989	0.3326	0.2940	0.9079
434	6	5.0879	0.2688	0.1855	0.3482	0.2869	0.9367
324	7	5.2182	0.2236	0.1231	0.3195	0.2348	0.9521
248	8	5.5807	0.1783	0.0620	0.2898	0.1882	0.9472

data restrictions, we don't have same number of repeated measurements by woman, which difficult the forecasting prediction evaluation over time as the number of woman dropped out of the sample over time increases quickly. With this in mind, we found RMSE values twice times higher than the same metric presented in the Table 1, showing that this models are still not working well for some women in the test group. Same conclusion can be drive to CCC and Pearson correlation. However, as the CCC can be viewed as $\mathrm{CCC} = \mathrm{r} \times \mathrm{C_b}$, where r represents a measure of precision while $\mathrm{C_b}$ a measure of accuracy[29]. Thus, we can conclude that our model has a high accuracy, which potentially growing up as the number of repeated measurements increases in the test data. The lower value found for precision in the test set reflects that the explanatory variables used into the model is still not enough to explain most of the variability in the data, suggesting that further variables such as polycystic ovary, daily diet, country, among other ones, could be explored into FitrWoman app in order to reduce the uncertainty and, consequently, improving the forecast for a data set with new measurements on women, and new women in added to the cluster.

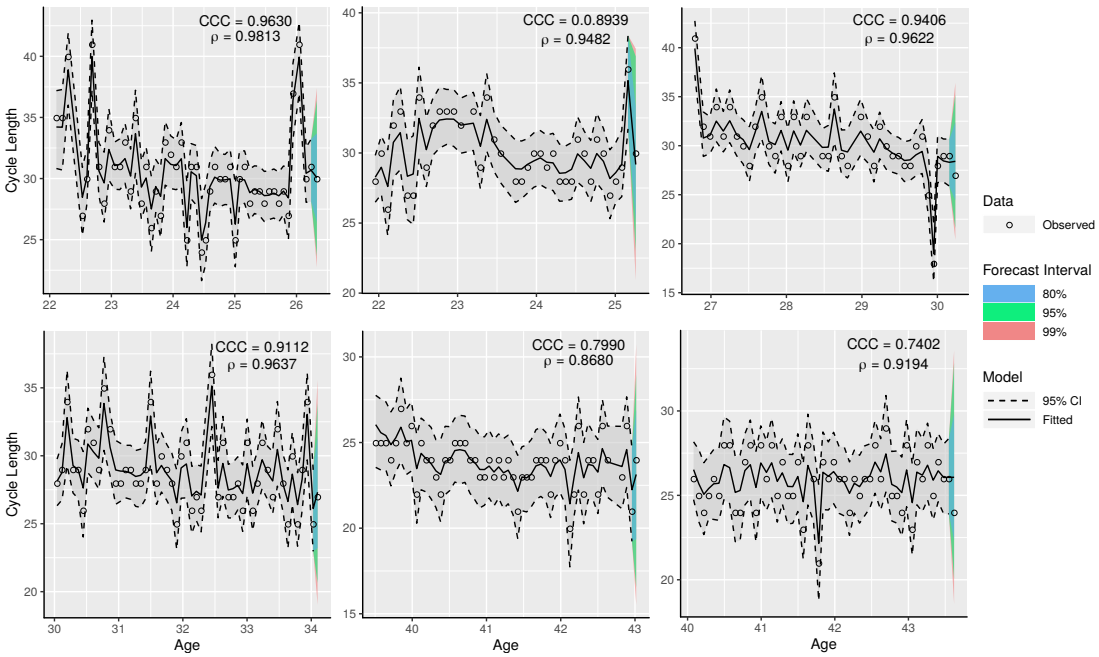


Figure 5. Age versus fitted menstrual cycle length for six women with more than 40 repeated measurements with addition of 95% credible interval (dashed line), 80%, 95%, and 99% forecast intervals for the next cycle, and observed menstrual cycle length as points. The estimated concordance correlation coefficient (CCC), and Pearson correlation coefficient (r) between fitted and observed values are described for each woman

Limitations

The limitation of this study is its observational study structure that depends on users include their information on the app and, consequently, does not provide good judgment about causality of reported symptoms on cycle length.

Conclusion

State-space models with a probability π to include a random effect at observational level in the random walk part are a useful approach for predicting menstrual cycle length, and it can be used to support female athlete wellness and optimize performance at all levels. In this sense, a random walk with overdispersion parameter and MA(1) model was selected to describe the complex dynamics menstrual cycle length over time, which resulted in high values of CCC and Pearson correlation between fitted and observed values by woman. Moreover, we demonstrated the importance in consider a overdispersed parameter to capture the variability of non-standard cycle by woman, which suggested that 26.81% $[24.14\%, 29.58\%]$ of cycles in our database are overdispersed, and the random walk standard deviation under a non-overdispersed cycle is $\sigma_\eta = 1.0530$ $[1.0060, 1.0526]$ days while under a overdispersed cycle it increases to $\sigma_w = 4.7386$ $[4.5379, 4.9492]$ days.

We also found a significant effect of heavy legs, increasing breathing, injury, stomach cramps, and flow amount in menstrual cycle length of female athletes whom are using the FitrWoman app. Although we found an accurate forecast prediction, some improvements in the data collection and model is still needed in order to increase the forecast precision such as consider an random effect of moving-average coefficient θ_0.

Methods

Data characteristics

The sample was comprised of female athletes using the FitrWoman app[30] who had given their consent for the use of their data for research purposes. The sample size contains data on 16,990 cycles collected from 2,178 women (Figure 7a), whose mean (sd) age was 33.89 (7.52) years old (range 14.98 - 47.10 years old); mean (sd) weight 62.73 (9.11) Kg (range 42.18-100.23 Kg); mean (sd) height 165.91 (6.87) cm (range 147.32-186.0 cm); and number of repeated measurements by woman ranging from 4-53 cycles. Although we have approximately 60% of non-information about woman's height and weight, the 95% quantile of the sample distribution based on 893 women are between 152.4 and

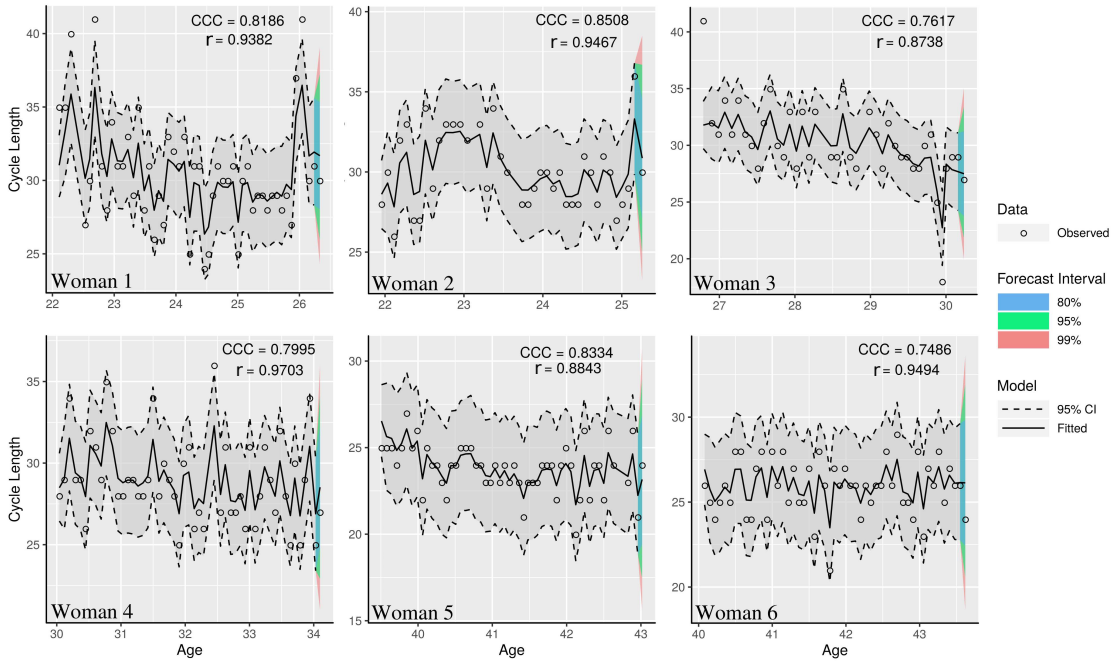


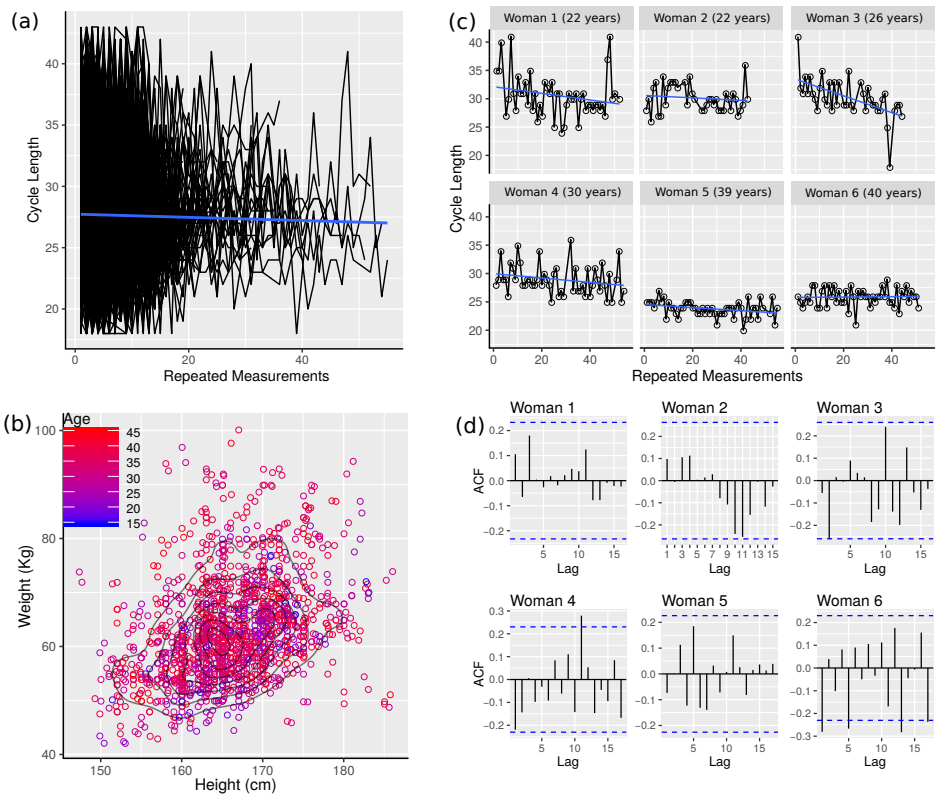
Figure 6. Age versus fitted menstrual cycle length for six women with more than 40 repeated measurements with addition of 95% credible interval (dashed line), 80%, 95%, and 99% forecast intervals for the next cycle, and observed menstrual cycle length as points using the model $y_{ij} = m_{ij} + \gamma_{ij} + c_{ij}$, where we just drop the term r_{ij} out. The estimated concordance correlation coefficient (CCC), and Pearson correlation coefficient (r) between fitted and observed values are described for each woman

180.0 cm, and 48 and 85 Kg for height and weight, respectively. Additionally, to give an idea about body structure of our cohort we build the bivariate density plot between weight and height given age, which is shown in the Figure 7b.

Menstrual cycle length is assumed to be normally distributed as we are working with standard cycles[12], where the shortest cycle length record was 18 days while the highest was 43 days, and the sample mean and variance are 27.66 and 3.54 days, respectively. As some women contributed with more than once sequence to the database, we decided to consider only the first sequence available because we don't no the reasons that caused this temporary dropout and the inclusion of the following sequences might bias the analysis, as also discussed by ref.[13].

Figure 7c shows six women profiles with a blue line representing a fitted mixed effects linear regression model. We can observe the inclusion of random intercept and slope plays an important rule as each woman can be affect by different non-observed explanatory variables. However, the conditional R^2 is equal 0.41, implying that the linear mixed-effects regression is a good approximation for some profiles, but not for all of them, diverging from the results presented by ref.[2], who fitted a simple linear regression and obtained an $R^2 = 0.99$. This may have happened because the amount of linear profiles observed by ref.[2] is suppressing the nonlinear profiles in their sample. In this sense and based on our sample, we have to account for woman's specific trend, temporal dependence among observations within woman, and variability across woman.

Figure 7c and 7d show that for some women a short cycle can be followed by a long one and vice-versa, suggesting the inclusion of a moving-average model. Furthermore, Figure 7d shows that some women have a positive autocorrelation while other ones have a negative autocorrelation suggesting the inclusion of an autoregressive moving-average model with individual random effect in the autocorrelation and/or in the moving-average coefficients. Finally, Figure 7e shows a table containing the marginal proportion of reported symptoms, where the partial conclusion is that most of they frequently not happen or are not reported. As consequence of possible misreported information, the effect of some symptoms on cycle length can be biased as the proportion of "No" presented in that table probably is confounded with true "No" and "Yes". However, it is important to note that the bias here is related to non-rejection of the null hypothesis (H_0 : there is no effect of a specific symptom) when the null hypothesis (H_0) is false (error type II). Despite the bias and power loss noted above, p values obtained with using statistical methods to data subject to random error or misclassification are still valid[31–33].



(e)

	Symptoms	No	Yes
1	Bloating	0.82	0.18
2	Constipation	0.94	0.06
3	Cravings	0.87	0.13
4	Diarrhea	0.91	0.09
5	Disturbed Sleep	0.82	0.18
6	Fatigue	0.81	0.19
7	Headache	0.87	0.13
8	Heavy legs	0.91	0.09
9	High temperature	0.92	0.08
10	Illness	0.96	0.04
11	Increased breathing	0.96	0.04
12	Injury	0.98	0.02
13	Irritability	0.86	0.14
14	Muscle ache	0.89	0.11
15	Nausea	0.94	0.06
16	Poor concentration	0.92	0.08
17	Stomach cramps	0.78	0.22
18	Stressed	0.89	0.11
19	Tender breasts	0.88	0.12
20	Weak	0.91	0.09
Levels of flow amount:			
21	Heavy	0.78	0.22
22	Medium	0.17	0.83
23	light	0.70	0.30
24	Spotting	0.81	0.19
25	None	0.89	0.11
Levels of training Intensity:			
26	Low	0.93	0.07
27	Medium	0.92	0.08
28	High	0.93	0.07

Figure 7. (a) Individual profile of 2178 women over time; (b) ; (c) Individual profile for six women from the data with addition of linear trend; (d) Bivariate density plot between weight and height given age; (e) Marginal proportion of symptoms reported, where the label "Yes" is related with "the event was reported at least once time" and "No" otherwise

Statistical analysis

Let Y_{ij} be a random variable, representing the length of menstrual cycle, of which y_{ij} is postulated to be a realization for the i-th woman, $i = 1, 2, \ldots, I$, at j-th menstrual cycle, $j = 1, 2, \ldots, J_i$. The main objective is the derivation of the one-step ahead predictive distribution given by

$$F_{iJ_i+1}\left(Y_{i,J_i+1}\right) = P\left(Y_{i,J_i+1} \leq y_{i,J_i+1} | y_{i1}, y_{i2} \ldots, y_{iJ}\right). \tag{1}$$

Consequently, we are interested in evaluating $F_{i,J_i+1}\left(Y_{i,J_i+1}\right)$ under a parsimonious parametric model, that is,

$$P\left(Y_{i,J_i+1} \leq y_{i,J_i+1} | y_{i1}, y_{i2} \ldots, y_{iJ}\right) = M_{i,J_i+1}\left(y_{i,J_i+1} | y_{i1}, y_{i2} \ldots, y_{iJ}, \boldsymbol{\Theta}^T\right),$$

where $M_{i,J_i+1}\left(y_{i,J_i+1}\right)$ is fully specified and $\boldsymbol{\Theta}$ is a vector of unknown fixed-effect and variance components parameters. Moreover, in order to accommodate the temporal correlation between repeated measures within woman and the variability across woman, we propose a selection procedure considering different error structures in the model.

Thus, random walk state-space model and mixed-effects state-space model are used and compared to account for both source of variation, incorporating a Bayesian approach for process forecasting to predict the duration and forecast interval, in days, of the next menstrual cycle. The computation of interval forecast is most appropriate in relation to point prediction because the last one is limited value as it only describe one of possible outcomes[34]. We assumed that women are independent among them; and menstrual cycles tends to decrease over a long period of time, being extremely related to woman's age[12,13]. In addition, we combine the Bayesian approach with forecasting process and covariate selection in the forecast, and a model validation procedure is made to demonstrate the effectiveness of this combination.

The individual model

The state-space formulation is known due to its flexibility to work with discrete response variable and temporal dependence among observations within woman while mixed-effects model can be used to account for between-woman

variability. As the observed event is the difference, in days, between the interval from the first day of one bleeding episode up to and including the day before the next bleeding episode, observed cycle lengths are discrete. Let Y_{ij} be a continuous random variable, where y_{ij} is a realization of Y_{ij} which represents the observed cycle length. Furthermore, let O_{ij} be a discrete random variable, where o_{ij} is a realization of O_{ij} which represents the cycle length in days of a continuous process, that is, $y_{ij} = o_{ij} + \varepsilon_{ij}$. As we have no way to estimate the error term ε_{ij} (observation process), we assumes here that $o_{ij} = \lfloor y_{ij} \rceil$ is a good approximation for y_{ij}, where $\lfloor . \rceil$ indicates rounding. Thus, the true non-observed continuous cycle length y_{ij} can be generated by the random walk state-space model:

$$
\begin{aligned}
y_{ij} &= m_{ij} + \gamma_{ij} + c_{ij} + r_{ij}, \\
m_{ij} &= m_{i,j-1} + \eta_{ij}, \text{ with } \eta_{ij} \sim N(0, \sigma_\eta^2), \\
\gamma_{ij} &= \phi\gamma_{i,j-1} + \theta\epsilon_{i,j-1} + \epsilon_{ij}, \text{ with } \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \\
c_{ij} &= \textstyle\sum_{k=1}^{K} \alpha_k C_{ijk} \\
r_{ij} &= \lambda_{ij} w_{ij},\ w_{ij} \sim N\left(0, \sigma_w^2\right),\ \lambda_{ij} \sim \text{Bernoulli}\left(\pi\right),\ \pi \sim \text{Uniform}\left(0,1\right)
\end{aligned}
\tag{2}
$$

where y_{ij} is the menstrual cycle length for the i-th woman at j-th cycle; m_{it} is a random walk model that allows a individual trend in the series with η_{ij} assumed to be normally distributed with mean 0 and variance σ_η^2. We specify a ARMA(1,1) model for γ_{it}, where ϕ is the autoregressive parameter; θ is the moving average parameter; and ϵ_{ij} is assumed to be normally distributed with mean 0 and variance σ_ϵ^2 (process error). Furthermore, c_{ij} has additional symptoms predictors (C_{ij}) that may be useful to understand the effect of them on cycle length, and, consequently, could be used in the forecasting process, where α_k represents k-th fixed effect parameter. Finally, r_{ij} is a random effect term responsible to account for extra-variability (overdispersion) of some menstrual cycle length measured on i-th woman at cycle j, which could be classified as outliers. Consequently, under model (2), y_{ij} has probability π of being an overdispersed menstrual cycle (non-standard) for the j-th cycle measured on i-woman, where its additional magnitude is given by r_{ij}, as shown in Figure 8. In this way, we can interpret m_{ij} as the trend for a

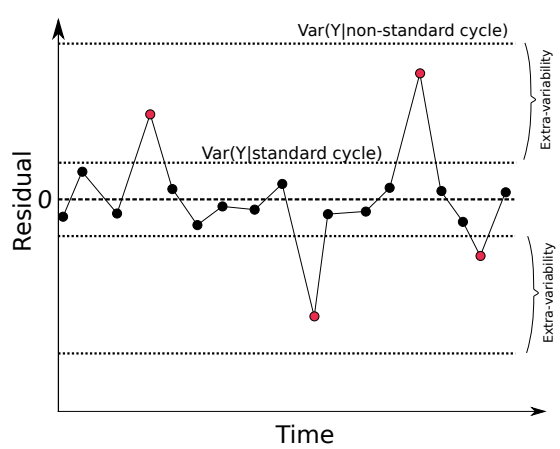


Figure 8. Representation of residuals ($\text{Residual}_{ij} = y_{ij} - \hat{m}_{ij}$) over time considering there are probability $1-\pi$ of Y be a standard cycle (non overdispersed) or π of Y to be a non-standard one (overdispersed), with $Var\left(Y|\text{standard cycle}\right) < Var\left(Y|\text{non-standard cycle}\right)$

standard cycle while $m_{ij} + r_{ij}$ could be interpreted as the trend for non-standard cycle, with r_{ij} working similarly as overdispersion parameter at observational level just for measures which induces extra-variability, as discussed in[35].

The state-space representation of the model (2) using the definition described by ref.[36] is given by

$$
\begin{aligned}
y_{ij} &= m_{ij} + \gamma_{ij} + \theta x_{ij} + c_{ij} + r_{ij}, \\
m_{ij} &= m_{i,j-1} + \eta_{ij}, \text{ with } \eta_{ij} \sim N(0, \sigma_\eta^2), \\
\gamma_{ij} &= \phi \gamma_{i,j-1} + \epsilon_{ij}, \text{ with } \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \\
x_{ij} &= \epsilon_{i,j-1}, \\
c_{ij} &= \textstyle\sum_{k=1}^{K} \alpha_k C_{ijk} \\
r_{ij} = \lambda_{ij} w_{ij},\ w_{ij} &\sim N\left(0, \sigma_w^2\right),\ \lambda_{ij} \sim \text{Bernoulli}\left(\pi\right),\ \pi \sim \text{Uniform}\left(0,1\right)
\end{aligned}
\tag{3}
$$

with initial value $m_{i1} \sim N\left(\beta_0, \sigma_\eta^2\right)$ for the local level model and $\gamma_{ij} = \sum_{t=0}^{j-1} \phi^t \epsilon_{i,j-t}$, with $j \geq 1$. The linear Gaussian state-space model defined by equation 3 are generated efficiently using the Kalman filter recursions[37].

On the other hand, assuming a specific linear regression by woman can be an important issue because the random intercept account for a natural variability relative to a uncertainty about how the non-observed variables affected the first observed menstrual cycle while a random slope assumes that each woman may have different menstrual cycle length trends relative to her age. To verify if the random walk model have the necessary flexibility to capture the different possible trends, we replace it by a linear mixed-effects model[13]. In that case, we rewritten $m_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\, Age_{ij}$, where β_0 and β_1 are the intercept and slope, respectively; b_{0i} and b_{1i} are random effects in the intercept and slope for the i-th woman at Age_{ij}, respectively, assumed to be

$$
\boldsymbol{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{G} = \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_{01}} \\ \sigma_{b_{01}} & \sigma_{b_1}^2 \end{bmatrix} \right),
$$

and Age_{ij} represents the woman's age.

Prior distribution

Bayesian analysis combines information from observed data with prior distribution for the model's parameters in order to generating a posterior distribution. Here the inverse-gamma(κ, κ) prior distributions are used for the random walk state-space model and variance components of mixed effect model as an attempt at non-informativeness within the conditionally conjugate family, with κ set to a low value such as 0.1^3:

$$
\sigma_\epsilon^{-2}, \sigma_\eta^{-2}, \sigma_w^{-2}, \sigma_{b_0}^{-2}, \sigma_{b_1}^{-2}, \sigma_\phi^{-2}, \sigma_\theta^{-2}, \sigma_\beta^{-2}, \sigma_{ar}^{-2} \sim \text{Gamma}\left(0.1^3, 0.1^3\right).
$$

We build a likelihood ratio test using a mixture of chi-square distributions and the 95% credible interval for the variance component σ_{b01} of mixed effect regression, whose result shows that the presence of correlations between the random effects in these models did not plays a crucial role and then can be removed from the model. Consequently, we can assume random effects are mutually independent.

Prior distribution for fixed effect parameters is given by $\beta_0 \sim N\left(\mu_{\beta_0}, \sigma_\beta^2\right)$, with $\mu_{\beta_0} \sim \text{Uniform}\left(24, 32\right)$; $\beta_1 \sim N\left(\mu_{\beta_1}, \sigma_\beta^2\right)$, with $\mu_{\beta_1} \sim \text{Uniform}\left(-2, 2\right)$; $\phi_0, \theta_0 \sim \mathcal{N}\left(\mu_{ar}, \sigma_{ar}^2\right)$, with $\mu_{ar} \sim \mathcal{N}\left(0, 100\right)$; and we assumed $\alpha_k \sim \mathcal{N}\left(0, 100\right)$, which is vague normal density prior. All assumptions were checked to make sure that results were not sensitive to specific choices of prior parameters.

Model selection and forecast

The model selection procedure adopted here is a balance between forecast accuracy and the Bayesian Information Criterion (BIC) used to compare candidate models, thus we split this procedure into three steps namely time trend component; autocorrelation component; and an additional linear predictor as a function of explanatory variables.

The first step is account for a possible trend, including the most appropriate error structure on the model, which in our case consists in build a random walk model or a linear mixed effect model (Figure 9). The second stage consists in including a temporal dependence among observations, as seems to be present for some women in our sample, where we considered the ARMA model as shown in the Figure 9. Finally, the last stage is the inclusion of predictors to account for the relationship between them and cycle length, where it can be done using the posterior distribution on the parameter α_k to select all those parameters that did not have the null value belongs to their 95%

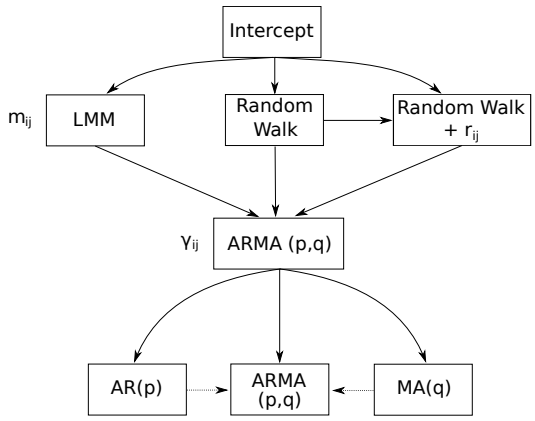


Figure 9. Stages 1 and 2 of the model selection procedure

LMM: linear mixed effect model; r_{ij}: overdispersion parameter at observational level; ARMA(p,q): Autoregressive moving average model of order p and q.

credibility interval, which contains the 95% most probable values. Finally, we checked back all model structures to make sure that we are using the most parsimonious model.

On the other hand, to evaluate the forecasting prediction we divided women randomly into two groups. The diagram of Figure 10 illustrates the series of training and test sets used to evaluate the forecasting accuracy within training set (group 1), and to cross-validation (group 2). The first group, comprising 2,178 women was divided again into training and test data, where the test data is composed by the last observation of each woman in order to access the forecasting accuracy for our training data (Figure 10a). Under other conditions, the second group, with 1,029 remaining women, was used to time series cross validation in a more sophisticated version of training and test sets, where the aim is verify the RMSE, CCC, Pearson correlation coefficient, as precision measure, and the bias corrector factor (C_b), as accuracy measure, over time when a new group of women is sequentially added to our training data (Figure 10b).

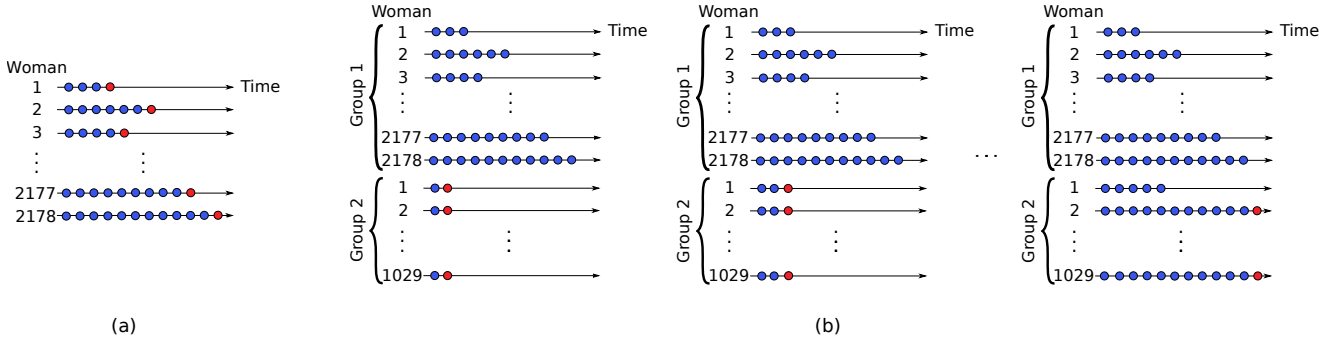


Figure 10. (a) One-step-ahead forecast to access the forecasting accuracy of training data; (b)time series cross validation for the new group (group 2) of women. Training data are represented by blue circles while test data by red ones.

The forecast error between an observed value and its forecasting was computed as

$$\epsilon_{i,J+h} = y_{i,J+h} - \hat{y}_{i,J+h|J}$$

where the train data is given by $\{y_{i1}, y_{i2}, \ldots, y_{iJ}\}$ and the test data is given by $\{y_{i,J+1}\}$ (one-step ahead predictive by woman), see Figure 10. Thus, we measure the forecast accuracy by summarising the root mean square error (RMSE), concordance correlation coefficient[29], and the Pearson correlation coefficient between the actual (test data) and fitted cycle length values.

Posterior computation

Markov Chain Monte Carlo (MCMC) method was used to generate samples from posterior distribution for the random walk and mixed-efects state-space models with Gibbs sampler algorithm[37], as this approach is widely used

MCMC to obtain the parameter estimate from the posterior distribution. Convergence of the MCMC algorithm is checked through multiple comparison of MCMC chains with dispersed starting points. In addition, the normality assumptions were checked using quantile-quantile plot with simulate envelop[38] in order to account the correlation between a given sample and the normal distribution and residual analysis was done using graphical analysis. The one-ahead predictive distribution of $F_{i,J_i+1}(Y_{i,J_i+1})$ was derived through draws from the posterior distribution. Consequently, the κ-step ahead predictive distribution can be obtained running sequentially the Kalman filter. All analysis was implemented in R including `runjags`[39], `coda`[40], `hnp`[38], and `ggplot2`[41] packages in our algorithm.

Data Availability

The data that support this study were made available by ORRECO. Due to privacy restrictions the data are not publicly available.

References

1. Regidor, P. A., Kaczmarczyk, M., Schiweck, E., Goeckenjan-Festag, M. & Alexander, H. Identification and prediction of the fertile window with a new web-based medical device using a vaginal biosensor for measuring the circadian and circamensual core body temperature. *Gynecol. Endocrinol.* **34**, 256–260 (2018). URL https://doi.org/10.1080/09513590.2017.1390737. DOI 10.1080/09513590.2017.1390737.
2. Bull, J. R. *et al.* Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. *npj Digit. Medicine* **2** (2019). URL http://dx.doi.org/10.1038/s41746-019-0152-7. DOI 10.1038/s41746-019-0152-7.
3. Symul, L., Wac, K., Hillard, P. & Salathé, M. Assessment of menstrual health status and evolution through mobile apps for fertility awareness. *npj Digit. Medicine* **2** (2019). URL http://dx.doi.org/10.1038/s41746-019-0139-4. DOI 10.1038/s41746-019-0139-4.
4. Ali, R., Gürtin, Z. B. & Harper, J. C. Do fertility tracking apps offer women useful information about their fertile window? *Reproductive BioMedicine Online* **00**, 1–10 (2020). URL https://doi.org/10.1016/j.rbmo.2020.09.005. DOI 10.1016/j.rbmo.2020.09.005.
5. Scherwitzl, E. B., Hirschberg, A. L. & Scherwitzl, R. Identification and prediction of the fertile window using NaturalCycles. *Eur. J. Contracept. Reproductive Heal. Care* **20**, 403–408 (2015). DOI 10.3109/13625187.2014.988210.
6. Schoene, R. B., Robertson, H. T. & Pierson, D. J. Respiratory drives and exercise in menstrual cycles of athletic and nonathletic women. *J. Appl. Physiol. Respir. Environ. Exerc. Physiol.* **50**, 1300–1305 (1981). DOI 10.1152/jappl.1981.50.6.1300.
7. Duane, M., Contreras, A., Jensen, E. T. & White, A. The performance of fertility awareness-based method apps marketed to avoid pregnancy. *J. Am. Board Fam. Medicine* **29**, 508–511 (2016). DOI 10.3122/jabfm.2016.04.160022.
8. Harlow, S. D. & Matanoski, G. M. The association between weight, physical activity, and stress and variation in the length of the menstrual cycle. *Am. J. Epidemiol.* **133**, 38–49 (1991). DOI 10.1093/oxfordjournals.aje.a115800.
9. Harlow, S. D. & Zeger, S. L. An application of longitudinal methods to the analysis of menstrual diary data. *J. Clin. Epidemiol.* **44**, 1015–1025 (1991).
10. Harlow, S. D., Lin, X. & Ho, M. J. Analysis of menstrual diary data across the reproductive life span Applicability of the bipartite model approach and the importance of within-woman variance. *J. Clin. Epidemiol.* **53**, 722–733 (2000). DOI 10.1016/S0895-4356(99)00202-4.
11. Li, K. *et al.* Characterizing physiological and symptomatic variation in menstrual cycles using self-tracked mobile-health data. *npj Digit. Medicine* **3**, 1–13 (2020). URL http://dx.doi.org/10.1038/s41746-020-0269-8. DOI 10.1038/s41746-020-0269-8. 1909.11211.
12. Guo, Y., Manatunga, A. K., Chen, S. & Marcus, M. Modeling menstrual cycle length using a mixture distribution. *Biostatistics* **7**, 100–114 (2006). DOI 10.1093/biostatistics/kxi043.
13. Bortot, P., Masarotto, G. & Scarpa, B. Sequential predictions of menstrual cycle lengths. *Biostatistics* **11**, 741–755 (2010). DOI 10.1093/biostatistics/kxq020.
14. Fukaya, K., Kawamori, A., Osada, Y., Kitazawa, M. & Ishiguro, M. The forecasting of menstruation based on a state-space modeling of basal body temperature time series. *Stat. Medicine* **36**, 3361–3379 (2017). DOI 10.1002/sim.7345. 1606.02536.

15. Lieberman, J. L., De Souza, M. J., Wagstaff, D. A. & Williams, N. I. Menstrual Disruption with Exercise Is Not Linked to an Energy Availability Threshold. *Medicine Sci. Sports Exerc.* **50**, 551–561 (2018). DOI 10.1249/MSS.0000000000001451.

16. Diaz, A., Laufer, M. R. & Breech, L. L. Menstruation in girls and adolescents: Using the menstrual cycle as a vital sign. *Pediatrics* **118**, 2245–2250 (2006). DOI 10.1542/peds.2006-2481.

17. Mumford, S. L. *et al.* The utility of menstrual cycle length as an indicator of cumulative hormonal exposure. *The J. clinical endocrinology metabolism* **97**, E1871–E1879 (2012). URL https://pubmed.ncbi.nlm.nih.gov/22837188https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3674299/. DOI 10.1210/jc.2012-1350.

18. Gizzo, S. *et al.* Menstrual cycle length: A surrogate measure of reproductive health capable of improving the accuracy of biochemical/sonographical ovarian reserve test in estimating the reproductive chances of women referred to ART. *Reproductive Biol. Endocrinol.* **13**, 1–11 (2015). DOI 10.1186/s12958-015-0024-1.

19. Mountjoy, M. *et al.* International Olympic Committee (IOC) Consensus statement on relative energy deficiency in sport (red-s): 2018 update. *Int. J. Sport Nutr. Exerc. Metab.* **28**, 316–331 (2018). DOI 10.1123/ijsnem.2018-0136.

20. Melin, A. K., Heikura, I. A., Tenforde, A. & Mountjoy, M. Energy availability in athletics: Health, performance, and physique. *Int. J. Sport Nutr. Exerc. Metab.* **29**, 152–164 (2019). DOI 10.1123/ijsnem.2018-0201.

21. Corbel, M. J., Tolari, F. & Yadava, V. K. Appropriate body-mass index for Asian populations and its implications. *The Lancet* **363**, 157–163 (2004). DOI 10.1016/S0140-6736(03)15268-3. arXiv:1011.1669v3.

22. Najmabadi, S. *et al.* Menstrual bleeding, cycle length, and follicular and luteal phase lengths in women without known subfertility: A pooled analysis of three cohorts. *Paediatr. Perinat. Epidemiol.* **34**, 318–327 (2020). DOI 10.1111/ppe.12644.

23. Treloar, A. E., Boynton, R. E., Behn, B. G. & Brown, B. W. Variation of the human menstrual cycle through reproductive life. *Int J Fertil.* **12**, 77–26 (1967).

24. Chiazze, L., Brayer, F. T., Macisco, J. J., Parker, M. P. & Duffy, B. J. The Length and Variability of the Human Menstrual Cycle. *JAMA* **203**, 377–380 (1968). DOI doi:10.1001/jama.1968.03140060001001.

25. Vollman, R. F. The menstrual cycle. *Major Probl Obstet Gynecol* **7**, 1–193 (1977).

26. Colombo, B. & Bassi, F. Studi in onore di Giampiero Landenna,. *Studi onore di Giampiero Landenna* 111–126 (1996).

27. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).

28. Montgomery, D., Peck, E. A. & ViningG G. *Introduction to linear regression analysis* (John Wiley & Sons, 2012), 5 edn.

29. Lin, L. I. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **45**, 255–268 (1989).

30. team, O. FitrWoman (2018). URL https://www.fitrwoman.com/.

31. Stefanski, L. A. The effects of measurement error on parameter estimation. *Biometrika* **72**, 583–592 (1985). DOI 10.1093/biomet/72.3.583.

32. Lagakos, W. S. Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. *Stat. Medicine* **7**, 257–274 (1988).

33. Buonaccorsi, J. P., Laake, P. & Veierod, M. B. On the effect of misclassification on bias of perfectly measured covariates in regression. *Biometrics* **61**, 831–836 (2005). DOI 10.1111/j.1541-0420.2005.00336.x.

34. Christoffersen, P. F. Evaluating Interval Forecasts. *Int. Econ. Rev.* **39**, 841–862 (1998).

35. Oliveira, T. d. P. & Moral, R. d. A. Global Short-Term Forecasting of Covid-19 Cases. *arXiv:2006.00111* 1–22 (2020). URL http://arxiv.org/abs/2006.00111. 2006.00111.

36. Brockwell, P. & Davis, R. *Introduction to Time Series and Forecasting* (Springer-Verlag, New York, 2002), 2 edn.

37. Carter, A. C. K. & Kohn, R. On Gibbs Sampling for State Space Models. *Biometrika* **81**, 541–553 (1994).

38. Moral, R. A., Hinde, J. & Demétrio, C. G. Half-normal plots and overdispersed models in R: The hnp package. *J. Stat. Softw.* **81** (2017). DOI 10.18637/jss.v081.i10.

39. Denwood, M. J. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *J. Stat. Softw.* **71** (2016). DOI 10.18637/jss.v071.i09.

40. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).

41. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2016), 2 edn.

Acknowledgements

The authors are grateful to National University of Ireland Galway and Orreco, that supported this research project. We extend our thanks for the Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289, co-funded by the European Regional Development Fund.

Author contributions statement

T.P.O. and J.N. conceived and implemented the modelling framework and wrote the manuscript with input from all coauthors. G.B. and C.P. designed the observational study and data collection. All the authors commented and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.N. or T.P.O.

Figures

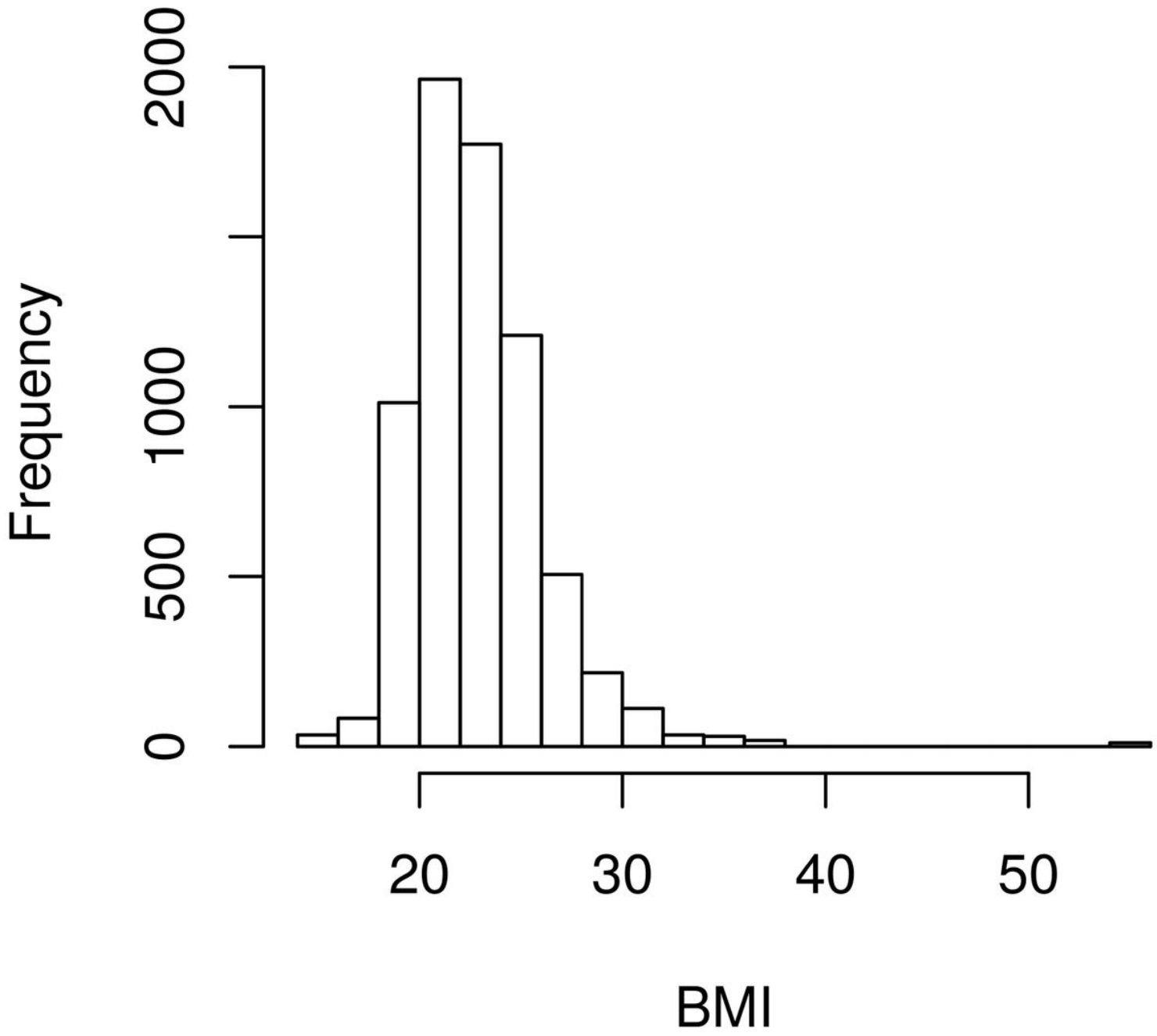


Figure 1

Histogram with BMI absolute frequency; and classification based on body mass index (BMI)

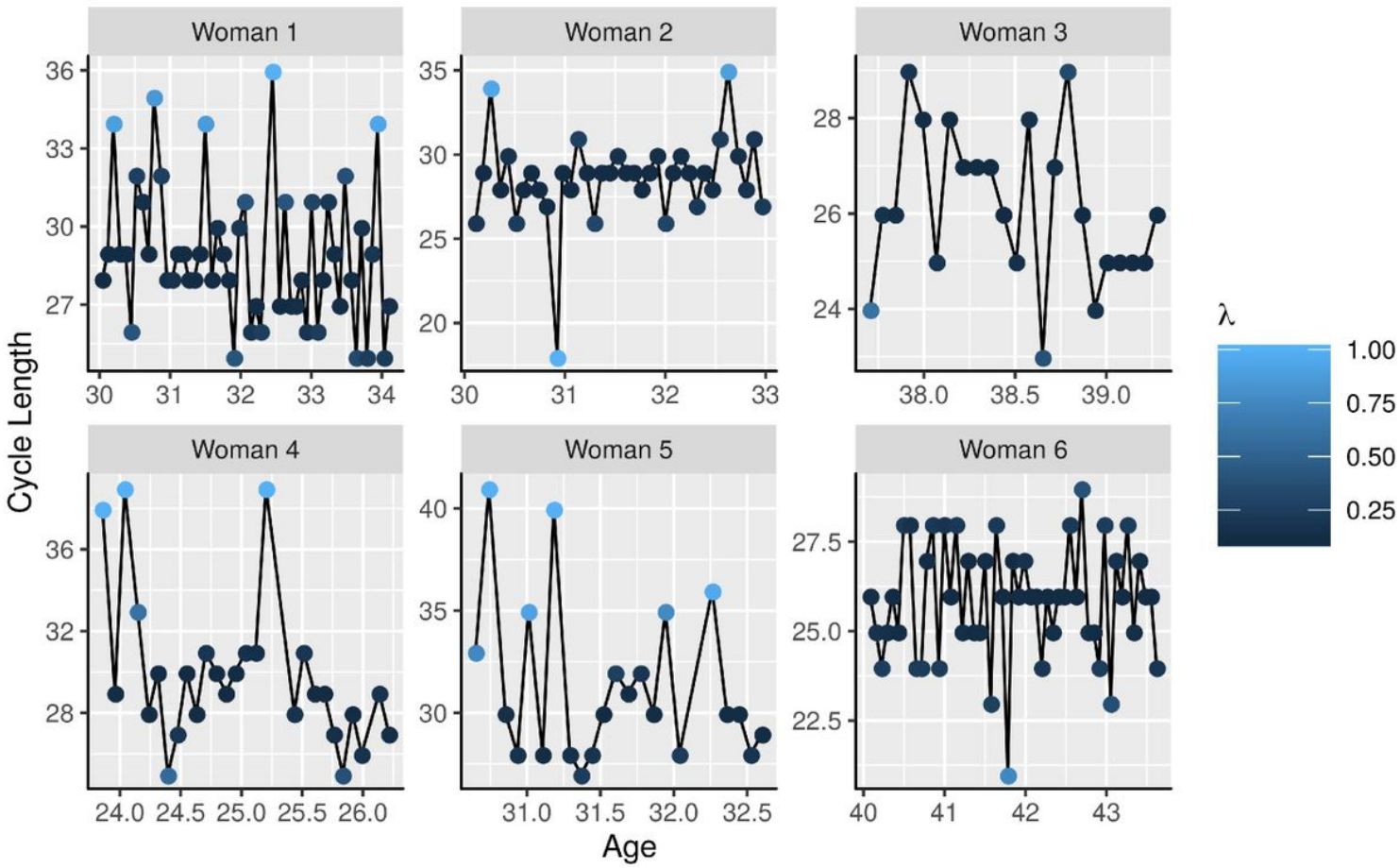


Figure 2

Example of six women profiles showing the probability that the proposed model considers an observation as overdispersed, where λ represents the probability of λij being equal 1 for a given observation

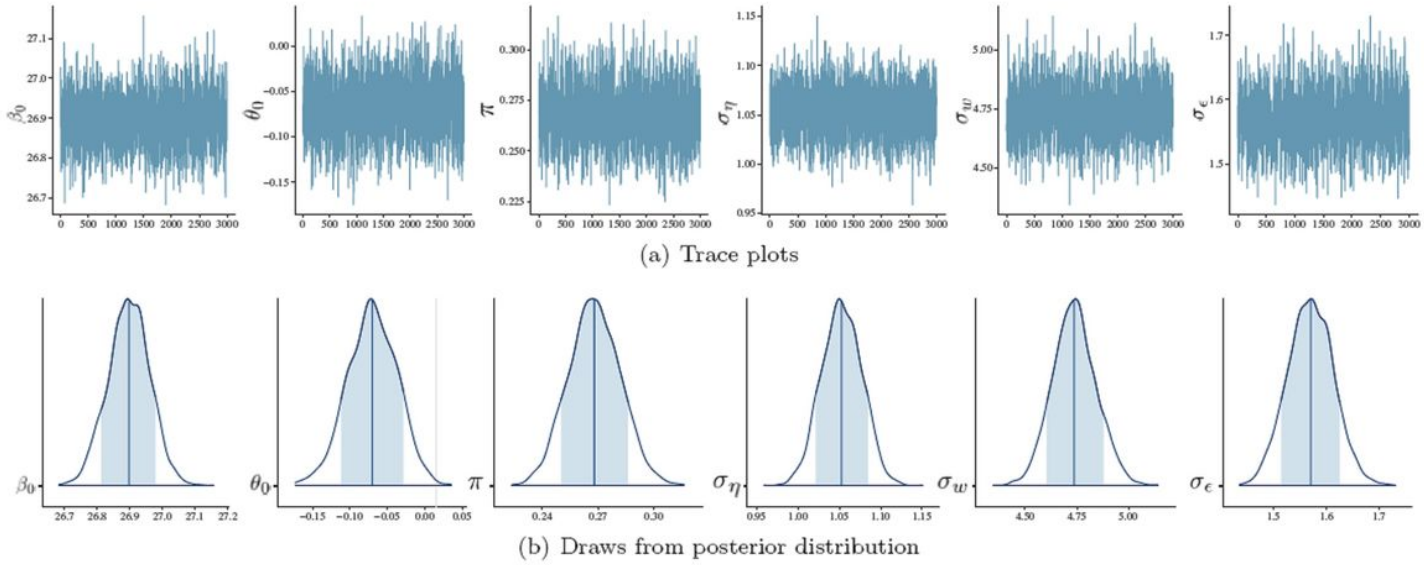
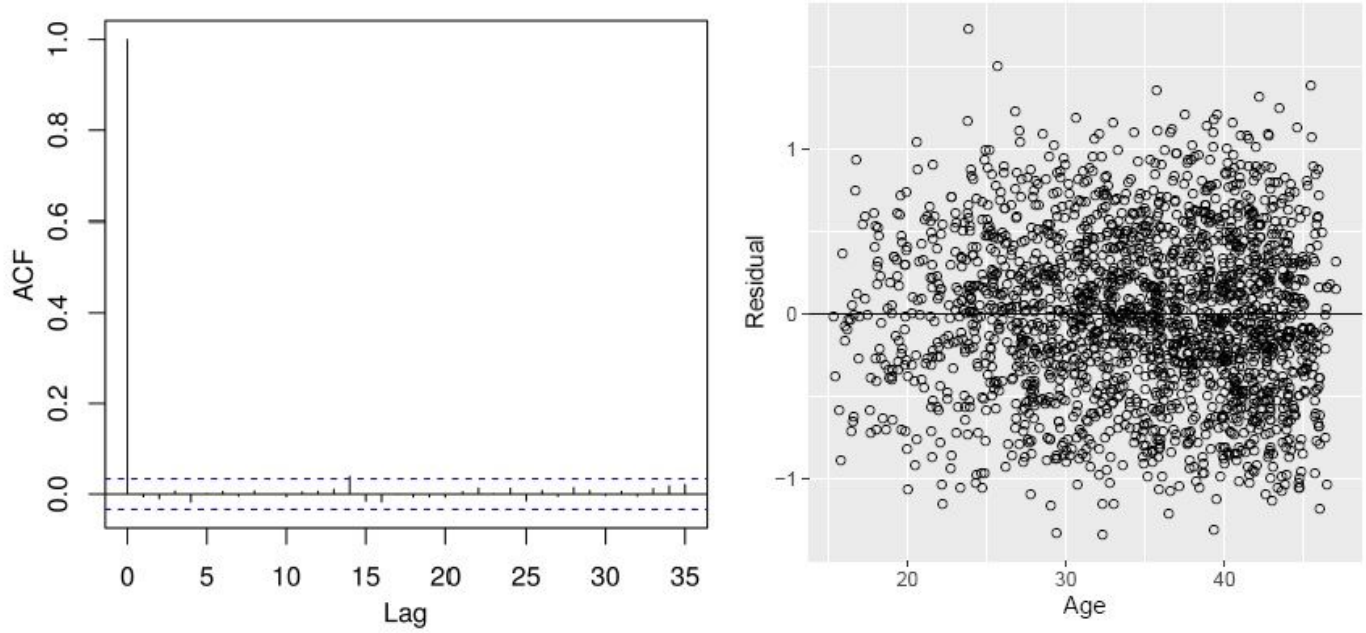


Figure 3

(a) Trace plots of Markov chains and (b) Markov chain Monte Carlo (MCMC) draws from the posterior distribution of the parameters β0, θ0, π, σn, σw, and σε, based on a sample of length 3000



(a) ACF (b) Residual vs. Age

Figure 4

(a) Residual autocorrelation plot, and (b) residual versus age

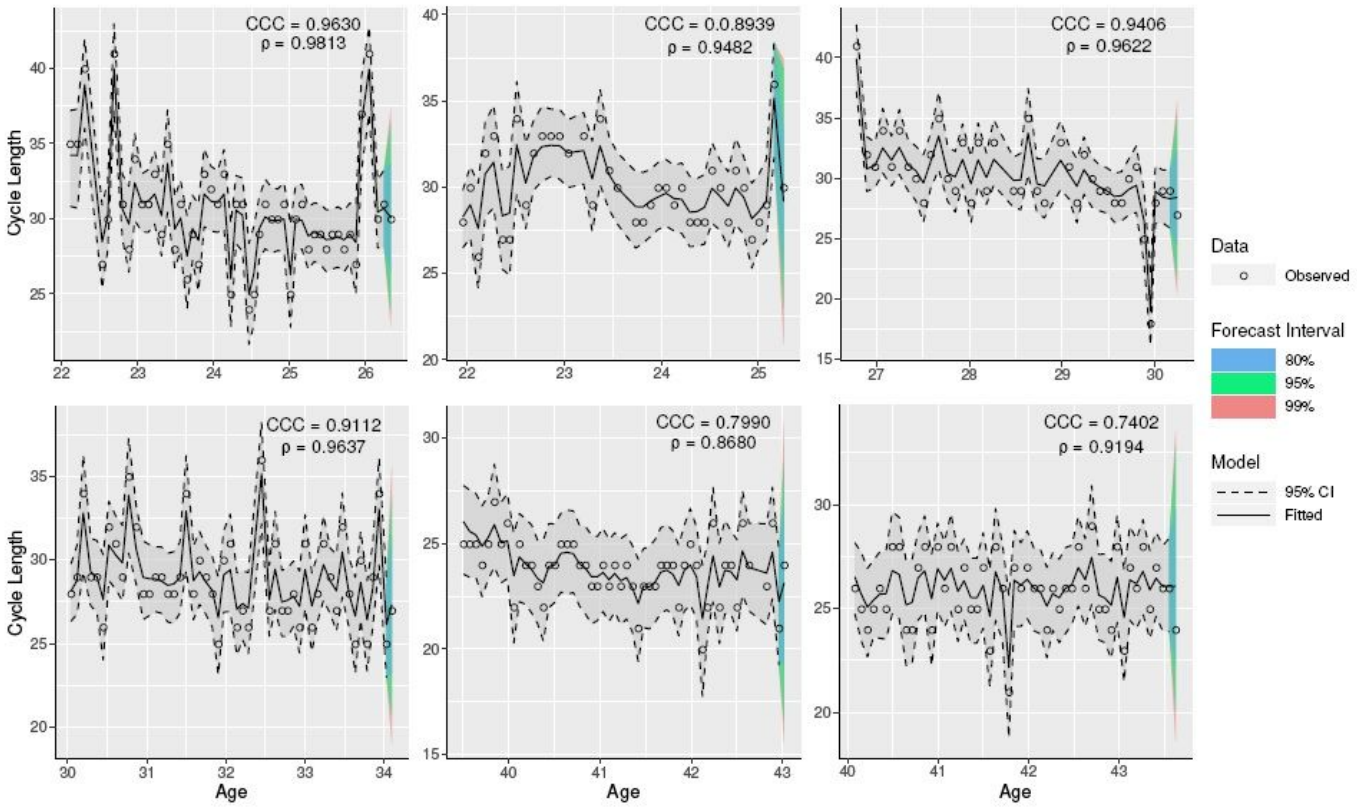


Figure 5

Age versus fitted menstrual cycle length for six women with more than 40 repeated measurements with addition of 95% credible interval (dashed line), 80%, 95%, and 99% forecast intervals for the next cycle, and observed menstrual cycle length as points. The estimated concordance correlation coefficient (CCC), and Pearson correlation coefficient (r) between fitted and observed values are described for each woman

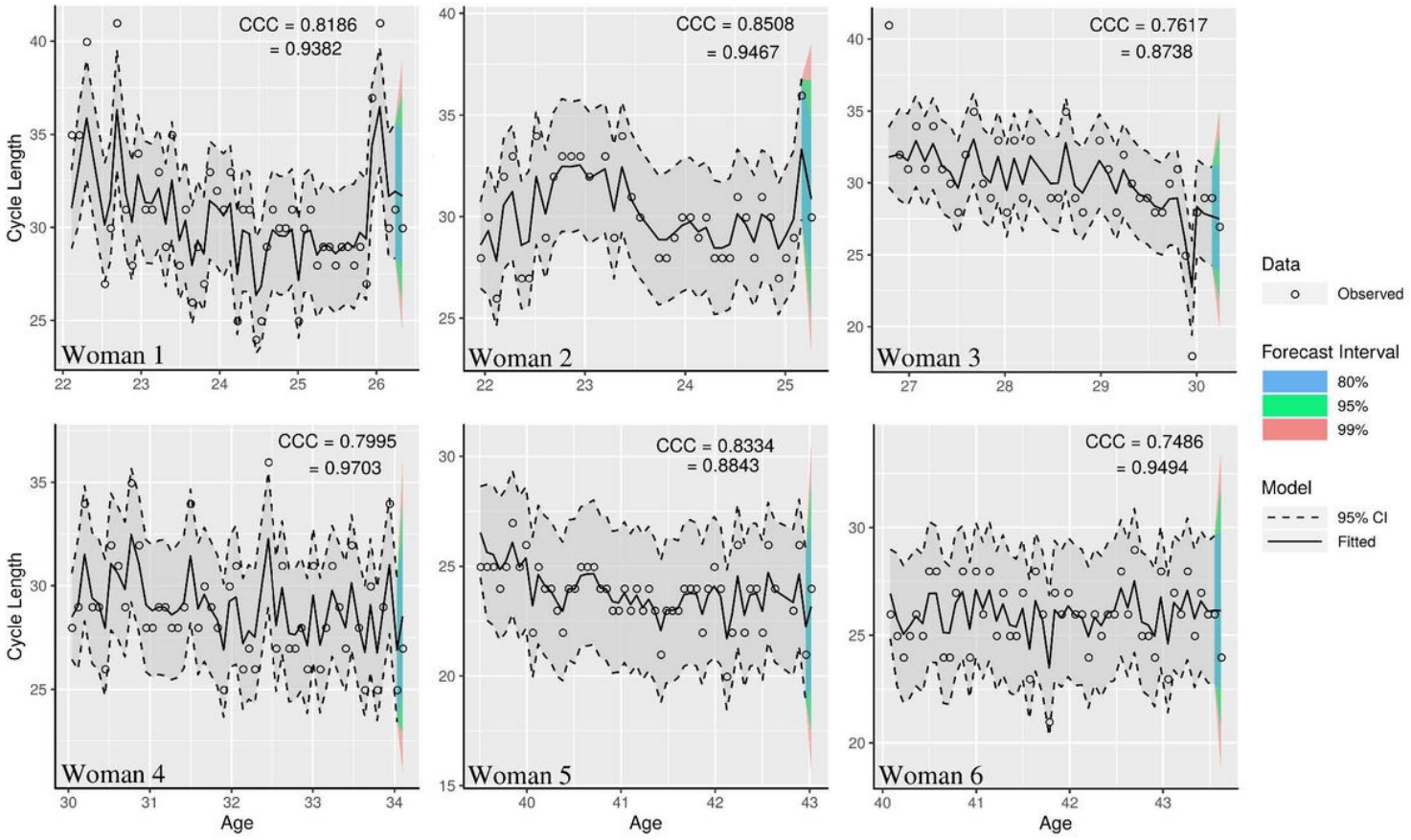
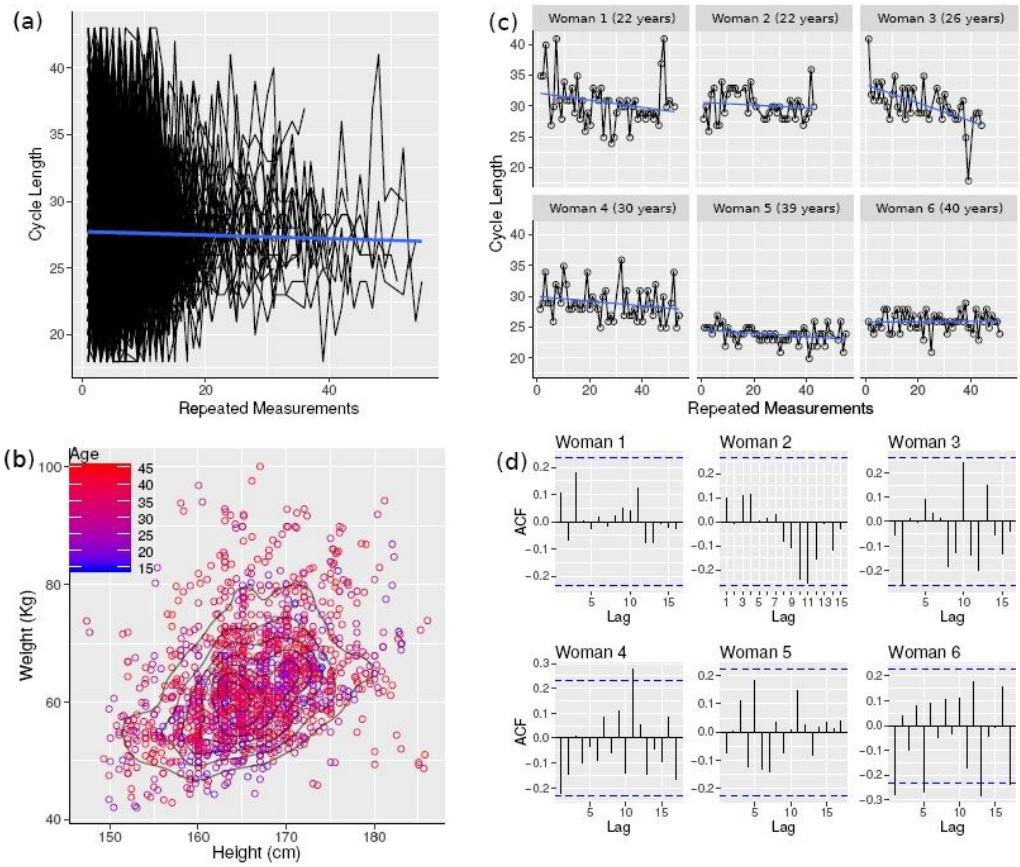


Figure 6

Age versus fitted menstrual cycle length for six women with more than 40 repeated measurements with addition of 95% credible interval (dashed line), 80%, 95%, and 99% forecast intervals for the next cycle, and observed menstrual cycle length as points using the model yij = mij + γij +cij , where we just drop the term rij out. The estimated concordance correlation coefficient (CCC), and Pearson correlation coefficient (r) between fitted and observed values are described for each woman



(e)

	Symptoms	No	Yes
1	Bloating	0.82	0.18
2	Constipation	0.94	0.06
3	Cravings	0.87	0.13
4	Diarrhea	0.91	0.09
5	Disturbed Sleep	0.82	0.18
6	Fatigue	0.81	0.19
7	Headache	0.87	0.13
8	Heavy legs	0.91	0.09
9	High temperature	0.92	0.08
10	Illness	0.96	0.04
11	Increased breathing	0.96	0.04
12	Injury	0.98	0.02
13	Irritability	0.86	0.14
14	Muscle ache	0.89	0.11
15	Nausea	0.94	0.06
16	Poor concentration	0.92	0.08
17	Stomach cramps	0.78	0.22
18	Stressed	0.89	0.11
19	Tender breasts	0.88	0.12
20	Weak	0.91	0.09
Levels of flow amount:			
21	Heavy	0.78	0.22
22	Medium	0.17	0.83
23	light	0.70	0.30
24	Spotting	0.81	0.19
25	None	0.89	0.11
Levels of training Intensity:			
26	Low	0.93	0.07
27	Medium	0.92	0.08
28	High	0.93	0.07

Figure 7

(a) Individual profile of 2178 women over time; (b) ; (c) Individual profile for six women from the data with addition of linear trend; (d) Bivariate density plot between weight and height given age; (e) Marginal proportion of symptoms reported, where the label "Yes" is related with "the event was reported at least once time" and "No" otherwise

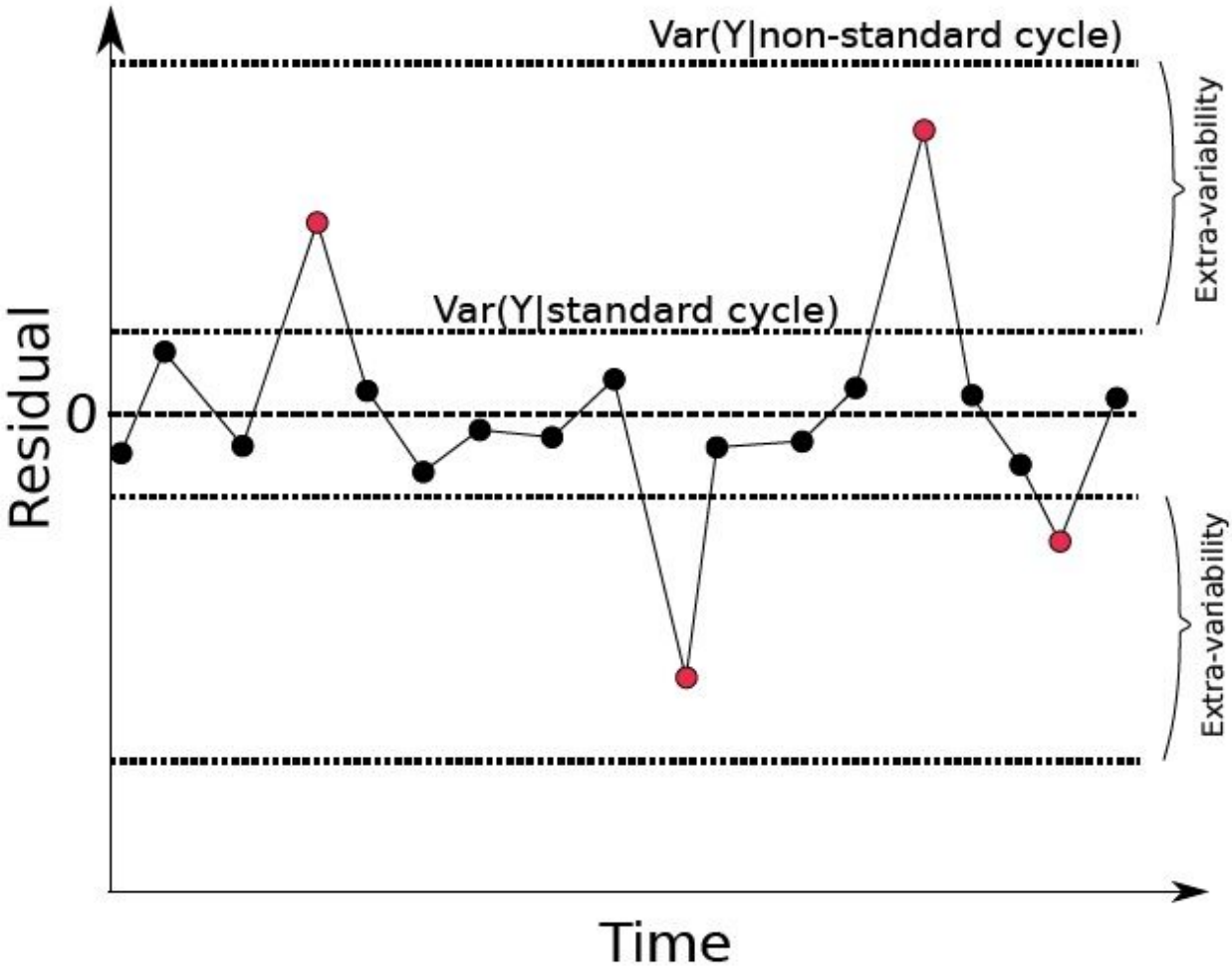


Figure 8

Representation of residuals (Residualij = yij − ˆmij) over time considering there are probability 1−π of Y be a standard cycle (non overdispersed) or π of Y to be a non-standard one (overdispersed), with V ar (Y |standard cycle) < V ar (Y |non-standard cycle)

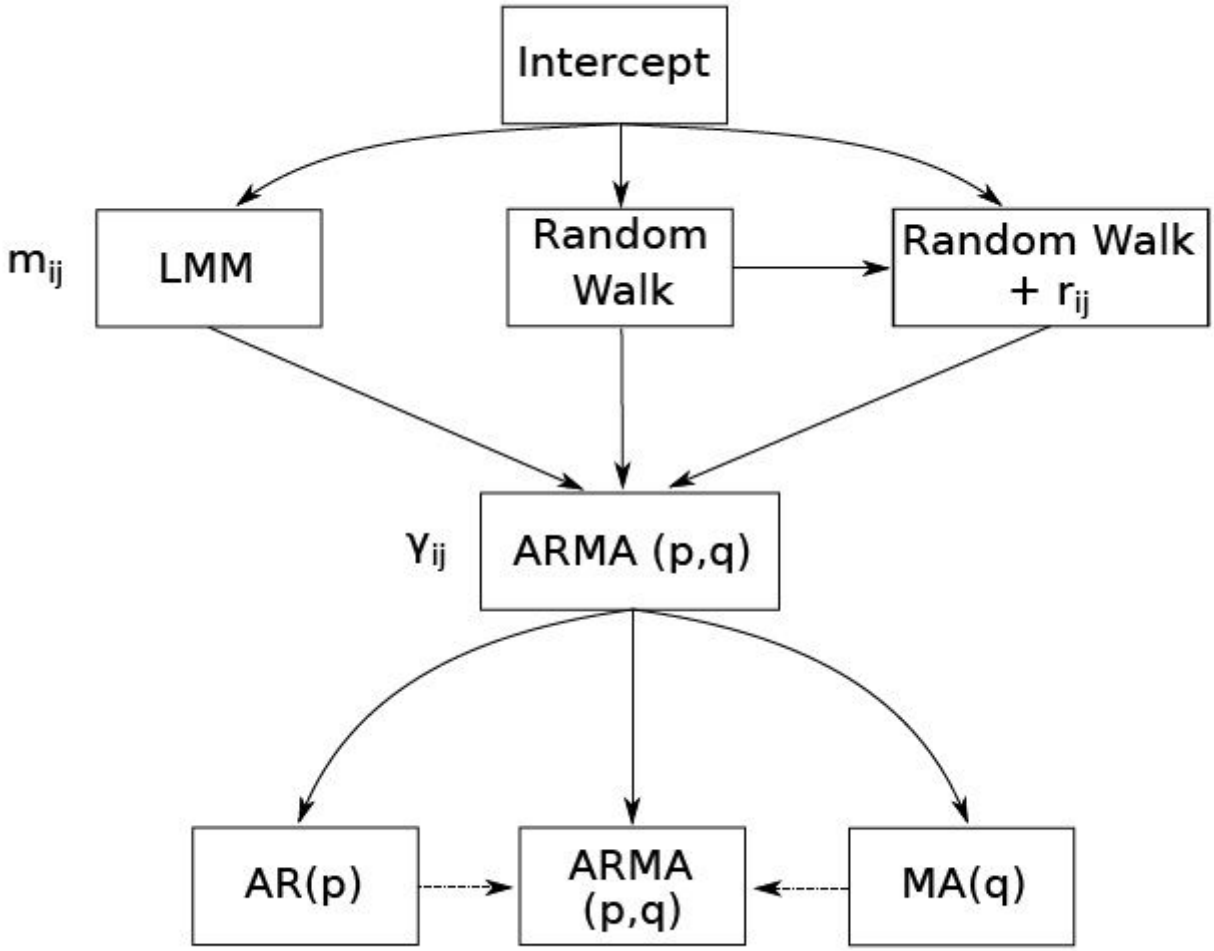


Figure 9

Stages 1 and 2 of the model selection procedure LMM: linear mixed effect model; rij : overdispersion parameter at observational level; ARMA(p,q): Autoregressive moving average model of order p and q.

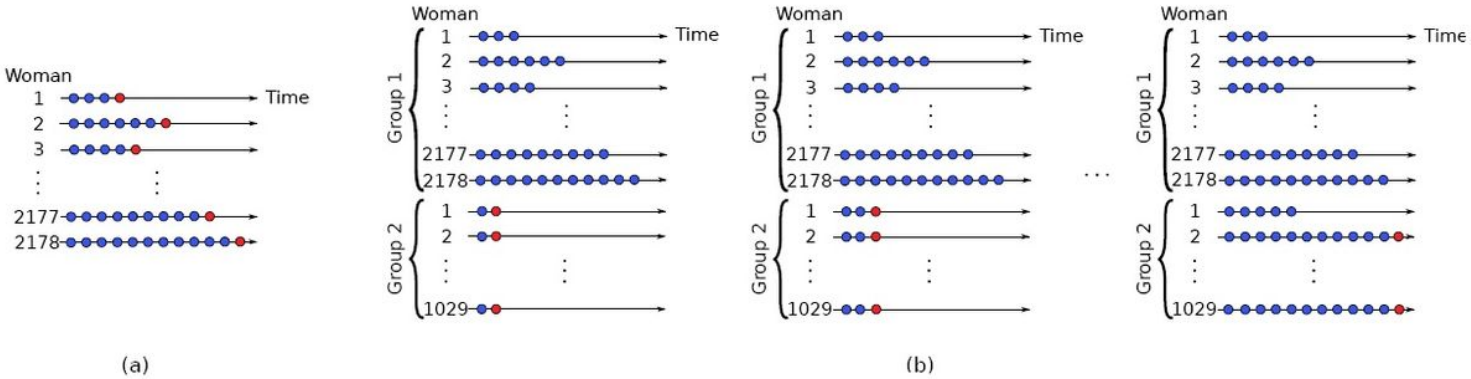


Figure 10

(a) One-step-ahead forecast to access the forecasting accuracy of training data; (b)time series cross validation for the new group (group 2) of women. Training data are represented by blue circles while test data by red ones.