Modeling Individual Cyclic Variation in Human Behavior

Emma Pierson
Stanford University
emmap1@cs.stanford.edu

Tim Althoff
Stanford University
althoff@cs.stanford.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

ABSTRACT

Cycles are fundamental to human health and behavior. Examples include mood cycles, circadian rhythms, and the menstrual cycle. However, modeling cycles in time series data is challenging because in most cases the cycles are not labeled or directly observed and need to be inferred from multidimensional measurements taken over time. Here, we present *Cyclic Hidden Markov Models* (CyHMMs) for detecting and modeling cycles in a collection of multidimensional heterogeneous time series data. In contrast to previous cycle modeling methods, CyHMMs deal with a number of challenges encountered in modeling real-world cycles: they can model multivariate data with both discrete and continuous dimensions; they explicitly model and are robust to missing data; and they can share information across individuals to accommodate variation both within and between individual time series.

Experiments on synthetic and real-world health-tracking data demonstrate that CyHMMs infer cycle lengths more accurately than existing methods, with 58% lower error on simulated data and 63% lower error on real-world data compared to the best-performing baseline. CyHMMs can also perform functions which baselines cannot: they can model the progression of individual features/symptoms over the course of the cycle, identify the most variable features, and cluster individual time series into groups with distinct characteristics. Applying CyHMMs to two real-world health-tracking datasets—of human menstrual cycle symptoms and physical activity tracking data—yields important insights including which symptoms to expect at each point during the cycle. We also find that people fall into several groups with distinct cycle patterns, and that these groups differ along dimensions not provided to the model. For example, by modeling missing data in the menstrual cycles dataset, we are able to discover a medically relevant group of birth control users even though information on birth control is not given to the model.

ACM Reference Format:
Emma Pierson, Tim Althoff, and Jure Leskovec. 2018. Modeling Individual Cyclic Variation in Human Behavior. In *WWW 2018: The 2018 Web Conference, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3178876.3186052

1 INTRODUCTION

Detecting and modeling cyclic patterns is important for understanding human health and behavior [27]. Cycles appear in domains as varied as psychology (*e.g.*, mood cycles [26] and cyclic mood disorders [22, 49]); physiology (*e.g.*, circadian rhythms and hormonal cycles [1, 14]); and biology (*e.g.*, the cell cycle [42]). Modeling such cycles enables beneficial health interventions: modeling mood cycles can predict symptoms and guide treatments [41], and modeling circadian cycles can predict fatigue and improve safety [1, 18].

In the cycles described above, a latent state progresses cyclically and influences observed data. Consider for concreteness the running example of hormone cycles. A person's hormonal state—that is, the levels of hormones in their bloodstream—is generally unobserved, but it progresses cyclically and can influence observed behavior [50]. For example, in women, a single hormone cycle consists of moving through a series of states—follicular, late follicular, midluteal, late luteal [7]—before returning to the follicular state. (Throughout this paper, completing a *cycle* means progressing through all latent states in a defined order before returning to the original latent state.) Modeling hormone cycles includes several problems of interest in human health: How does one compute the length of a cycle for a particular person? Can we cluster individuals into groups with distinct cycle dynamics, who may benefit from different medical interventions? How do features/symptoms vary throughout the cycle, since features that increase at the same point in the cycle may have the same physiological cause? (We use *feature* to refer to a single observed dimension of a multivariate time series: in the hormone cycles case, a binary feature might be whether the person reported negative mood on each day, and a second feature might be whether they reported pain).

Modeling cycles presents a number of challenges. The cycle state for each individual at each timestep is generally unobserved: a person's cyclic hormonal state may influence the level of positive affect they express on social media, but social media data is rarely linked to data on hormone state. In addition, cyclic dynamics vary across individuals—for example, individuals may experience cycles of slightly different lengths—and vary over time within a single individual as well. Further, real-world time series data is multivariate with discrete as well as continuous features and missing data.

Numerous methods have been developed to model cyclic patterns, including Fourier transforms, autocorrelation, data mining techniques, and computational biology algorithms [5, 9, 13, 15, 27, 40, 51, 63]. However, these techniques fail to address one or more of the challenges described above and, because they are not generative models, lack the ability to fully model the cycle: for example, they are not designed to model how each feature varies throughout the cycle or cluster individuals into distinct groups. In addition, previous work has been limited by the lack of ground-truth data on the true latent cycle state, which has prevented quantitative evaluation and comparisons between methods. That is, if an unsupervised method claims that cycles have a certain length or a certain feature or symptom varies particularly dramatically over the course of the cycle, how do we assess whether those inferences are accurate?

This work. Here we present a new method, *Cyclic Hidden Markov Models (CyHMMs)*, which models cycles and addresses the challenges described above. CyHMMs take as input a collection of

WWW 2018, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5639-8/18/04.
https://doi.org/10.1145/3178876.3186052

multivariate time series, with one time series for every individual in a population; time series can have missing data and discrete or continuous values. For each individual and each timestep, CyHMMs infer a discrete latent cycle state from the observed time series, using the fact that features vary dynamically based on the latent cycle state. Given these inferred cycle states, CyHMMs can then (1) compute the *cycle length* as the time it takes an individual to return to the same latent state, (2) predict *feature trajectories* by using the fitted generative model to predict how each feature will change over the course of the cycle, (3) identify the *most variable features*, and (4) *cluster* individuals into groups with distinct cycle dynamics.

We evaluate CyHMMs on both simulated data and two real-world cyclic datasets—of human menstrual cycle data, and activity tracking data—where we have knowledge of the true cycle state. Using simulated data with known parameters, we show that CyHMMs recover the cycle length 58% more accurately than the best performing baseline. On real-world data, CyHMMs recover the cycle length 63% more accurately than the best performing baseline. Furthermore, we show that CyHMMs can accurately infer other fundamental aspects of the cycle which existing methods cannot:

- *Feature trajectories:* CyHMMs can infer how each feature or symptom fluctuates over the course of the cycle. This is important for understanding an individual's likely time course through the cycle, and for understanding which features tend to peak during the same cycle state, potentially indicating a common cause.
- *Feature variability:* CyHMMs can infer which features are most variable. This is important for understanding which features are affected by the cycle, and which remain relatively constant.
- *Clustering:* CyHMMs can partition individuals into groups which correlate with true population heterogeneity. This is important because not all individuals have the same cycle dynamics, and so different individuals may benefit from different interventions.

Our inferences on two real-world datasets further demonstrate the utility of cycle modeling in the important human health domains of activity tracking, where we show we can accurately recover weekly sleep cycles, and menstrual cycle tracking, where we show we can accurately recover the 28-day cycle. For example, in the menstrual cycle data, CyHMMs identify a subpopulation of users who are much more likely to be taking hormonal birth control—even though no information on birth control is provided to the model. Identifying such subpopulations is essential for accurately characterizing cycles (since women taking birth control experience different menstrual cycle dynamics [45, 48]) and to designing medical treatments (since women taking birth control require different pharmacologic interventions [30] and have different risks of hormone-related cancers [44, 46]). We emphasize that while in this paper we analyze datasets where cycle states are known in order to prove that CyHMMs work, CyHMMs are generally applicable to time series where cyclic dynamics are suspected but cycle states are not known. Our CyHMM implementation is publicly available.

2 RELATED WORK

Our work draws on an extensive literature of methods developed for detecting and modeling cyclic patterns; we briefly describe major approaches. *Fourier transforms* [9] express a time-dependent signal as a sum of sinusoids or complex exponentials; the most significant frequencies in the signal can be extracted from the amplitudes of the Fourier coefficients. While Fourier transforms are a classic technique for univariate time series, they are not easily adapted to data which is multivariate, missing, or discrete. Further, because they are not generative models, they cannot easily be used to predict how features will vary over the course of the cycle. It is also not clear how to apply them to a population of individuals whose time series are different but related. Fourier transforms also assume that the cycle length in each signal remains constant, which is not true for individuals whose cycle lengths vary from cycle to cycle. (Wavelet transforms [15] overcome this last deficiency but not the other ones.) Another common technique for detecting periodicity, *autocorrelation*, computes the correlation between a time series and a lagged copy of itself [51]. It is mathematically related to the Fourier transform and suffers from similar drawbacks.

More recent *data mining techniques* find "partial periodic patterns" that occur repeatedly in time series, sometimes with imperfect regularity [5, 6, 12, 25, 27–29, 32, 34, 39, 40, 61, 62]. These methods represent discrete time series as strings and search for recurring patterns. Han et al. [27, 28] introduce the concept of partial periodicity; numerous related approaches have been developed, including Ma and Hellerstein's algorithm for detecting periodic events with unknown periods [40] and methods for detecting frequently recurring patterns [25, 29, 47]. These models do not address several challenges addressed in this work. They are not full generative models and do not recover all the cycle parameters in which we are interested; they cannot, for example, predict how observed features will vary over the course of the cycle. Second, we seek a method applicable to *both* discrete and continuous data, and string-based methods are inherently discrete.

There is also a relevant literature in computational biology which searches for circadian rhythms, often in genomic data [16, 33, 60, 63], by fitting a periodic model to individual genes. These models have several drawbacks for the real-world datasets we consider: they are designed for data which is much higher-dimensional than most human activity datasets; they are designed for continuous data, and may not work well on discrete data; they consider each feature individually rather than sharing information across features; and they do not easily allow sharing of information across individuals.

Because our method infers a hidden state at each timestep, it is also inspired by previous approaches which use latent states—and hidden Markov models specifically—to model time series. Such latent state approaches have been applied to model time series in domains such as speech recognition [52], healthcare [58] and sports [35]. These approaches have been successful because multivariate time series in many domains are often generated by low-dimensional hidden states, motivating our application to cyclic time series. Our work also draws on ideas from hidden semi-Markov models (HSMMs) and explicit duration hidden Markov models (EDHMMs) [21, 37, 65] which develop formalisms for modeling latent states with non-geometric duration distributions that are commonly found in real-world data [1, 36].

3 TASK DESCRIPTION

We assume the data consist of N multivariate time series, one for each individual i in a population. Throughout, we use *feature* to refer to a single dimension of a multivariate time series. Each time series, $X^{(i)}$, is a $T^{(i)} \times K$ matrix, where $T^{(i)}$ is the number of

timesteps in the time series and K is the number of features. $X_{tk}^{(i)}$ denotes the value in the i-th time series at the t-th timestep of the k-th feature. Features may be binary or continuous and may have missing data. Our task is to infer basic cycle characteristics:

- *Cycle length:* both for each individual and the whole population.
- *Feature progression:* how each feature varies over the course of the cycle.
- *Feature variability:* which features vary the most over the course of the cycle and which remain relatively constant.
- *Clustering:* whether the population can be divided into groups with distinct cycle characteristics.

4 PROPOSED MODEL

We propose a new class of generative models for cyclical time series data: *Cyclic Hidden Markov Models* (CyHMMs). As in a standard HMM, at each timestep an individual is in one of J latent states Z_j, indexed by j, and emits an observed feature vector drawn from a distribution specific to that latent state. However, CyHMMs differ from standard HMMs in two ways. First, they capture cyclicity by placing constraints on how individuals can transition between states: each individual must progress through states $Z_1, Z_2, \ldots$ in the same order (although they can begin in any start state) and after reaching the final state Z_J must return to Z_1 to begin the cycle again. An individual thus completes a *cycle* when they pass through all latent states and return to the same latent state. We constrain latent state progression in this way because it naturally encodes how the latent state in real-world cycles progresses cyclically and because it makes the model interpretable. If the latent state transitions were unconstrained it would not be clear how to define cycles.

The second difference between our model and an ordinary HMM is that, when an individual enters a new state Z_j, they also draw a number of timesteps—a *duration*—to remain in that state before transitioning to the next state, an idea from hidden semi-Markov models [65] (HSMMs) and explicit duration hidden Markov models [21, 37] (EDHMMs). In ordinary HMMs, the time spent in a state is drawn from a geometric distribution [17], but as we discuss below, this parameterization is a poor approximation when modeling many real-world cycles; drawing state durations from arbitrary duration distributions allows for more flexible and realistic cycle modeling. Figure 1 illustrates a path an individual might take through hidden states over the course of a single cycle. Each state Z_j is split into substates indexed by d, where d indicates the amount of time remaining in that state. Substate Z_{jd} progresses deterministically to $Z_{j(d-1)}$ unless $d = 0$, in which case it can progress to any substate of state $j + 1$ with probabilities drawn from $\mathcal{D}_{j+1}$, where $\mathcal{D}$ is the duration distribution. (If $j = J$, the model progresses instead to Z_1, completing the cycle). All substates of state Z_j share the same emission distribution $\mathcal{E}_j$, and all transitions are either deterministic or parameterized by the duration distributions $\mathcal{D}_j$, so the model's emissions and transitions are completely parameterized by $\mathcal{E}_j$ and $\mathcal{D}_j$[1]. Expressing the model in this way is a commonly used computational technique [65], facilitating fast parameter fitting through optimized HMM packages [54].

[1]While in theory some duration distributions can allow an infinite number of substates for each state, in practice duration distributions have negligible mass beyond a certain maximum duration d_{max}, so the number of substates can be safely truncated.

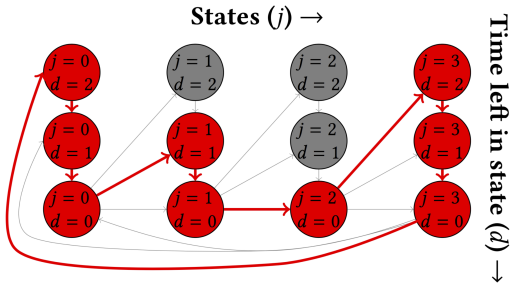

Figure 1: Latent states in a CyHMM with J = 4 states and maximum state duration $d_{max} = 2$. The red circles and arrows illustrate one possible path through the CyHMM over the course of a single cycle. The j-th latent state counts down until the remaining duration in the state, d, is zero and then transitions to the next latent state. All substates with equal j have the same emission parameters.

4.1 Model specification

Specifying a CyHMM requires specifying the emission distributions $\mathcal{E}$, the duration distributions $\mathcal{D}$, and the number of states J. We now describe and motivate our distributional choices. We find in our experiments that the distributions described below are sufficiently flexible to capture both simulated and real-world data, although our framework extends to other distributions.

Emission distributions. Human activity time series have frequent missing measurements caused, for example, by a user forgetting to track some event on a given day. We show in our real-world case studies (Section 6), that the probability of missing data fluctuates over the course of the cycle and is important to model. Consequently, we use emission distributions which allow for missing data. We model the data using a mixture where each observation has some probability of being missing and is otherwise drawn from a Gaussian distribution (for continuous features) or a Bernoulli distribution (for binary features)[2].

Continuous features: if a user is in the j-th latent state at timestep t, we assume their value for the k-th feature, X_{tk}, is drawn as follows:

$$X_{tk} = \begin{cases} \text{unobserved} & \text{if Bernoulli}(p_{jk}^{\text{obs}}) = 0 \\ v_{tk} \sim \mathcal{N}(\mu_{jk}, \sigma_{jk}^2) & \text{otherwise} \end{cases}$$

Hence, X_{tk} is unobserved with some probability $1 - p_{jk}^{\text{obs}}$ and otherwise drawn from a Gaussian whose parameters are specific to that state and feature. The parameters for the j-th latent state and k-th feature are μ_{jk}, σ_{jk}, and p_{jk}^{obs}. Each feature is independent conditional on the hidden state: the emission probability for all features is the product of the emission probabilities for each feature.

Binary features: Binary features in real-world time series introduce a subtlety: if all features are zero, it is unclear whether they are true zeros or missing data. For example, if a person on a negative mood-tracking app records no negative mood symptoms, it is unclear whether they were content or whether they merely neglected to log their negative mood. (With continuous features, it is generally more obvious when data is missing, since zero values are out of

[2]Categorical features could be modeled using a simple extension of our Bernoulli model; as there are no categorical features in our real-world datasets, we do not consider them here.

range). Thus, we assume the following data-generating process: with some probability at each timestep, the person does not bother to log any features and data for all features is missing. Otherwise, they log at least one feature, and we assume they did not experience the features they did not log. Hence, if a person is in the j-th latent state at the t-th timestep, observed data is drawn as follows:

$$I_t^{\text{logged anything}} \sim \text{Bernoulli}(p_j^{\text{obs}})$$

$$X_{tk} = \begin{cases} \text{unobserved} & \text{if } I_t^{\text{logged anything}} = 0 \\ v_{tk} \sim \text{Bernoulli}(E_{jk}) & \text{otherwise} \end{cases}$$

The parameters of the emission distribution are p_j^{obs} and E_{jk}.

Duration distributions. For ordinary HMMs, the time spent in each state follows a geometric distribution. However, this parameterization does not describe many real-world cycles. For example, the length of the luteal phase in the human menstrual cycle is better described by a normal distribution [36]. In particular, the monotonicity of the geometric probability mass function makes it a poor approximation to many real-world distributions. Instead, we use Poisson distributions for the duration distributions $\mathcal{D}_j$, so the duration distribution for the j-th latent state is parameterized by a single rate parameter λ_j. We use Poisson distributions because they are frequently used in HSMMs, they are fast to fit and easy to interpret, and we show empirically that they provide a better fit to real-world cycles than does an implementation with a geometric distribution. Our framework extends to other distributions.

Choosing the number of hidden states. J can be chosen either based on prior knowledge about the cycle being modeled (as we do in our case studies) or using cross-validated log likelihood [11].

4.2 Model fitting

The optimization procedure is an expectation-maximization algorithm often used with HMMs; we outline it briefly.

- **E-step**: Use the Baum-Welch algorithm [8] to infer the probability of being in a given latent state at a given timestep, $p_j^{(t)}$, and the expected number of transitions between each pair of substates Z_{jd} and $Z_{j'd'}$. Because of the model structure (Figure 1), the expected number of transitions is trivial for all start substates with $d \neq 0$ (since the substate must simply count down to the next substate). Denote by $C_{jd}^{(i)}$ the expected number of transitions in time series i into substate Z_{jd} from the prior substate $Z_{(j-1)0}$.
- **M-step**: Update the parameters of the duration distribution $\mathcal{D}_j$ and the emission distribution $\mathcal{E}_j$ for each latent state.
  - Updating $\mathcal{D}_j$: In the case of the Poisson distribution, each state has a single rate parameter λ_j which is estimated as the mean expected duration in that state:

$$\lambda_j = \frac{\sum_{i=1}^{N}\sum_{d=1}^{D} C_{jd}^{(i)} \cdot d}{\sum_{i=1}^{N}\sum_{d=1}^{D} C_{jd}^{(i)}}$$

    We set the probability of beginning in each substate Z_{jd} to be $f_j(d)$, where f_j is the duration probability mass function for the j-th latent state.
  - Updating $\mathcal{E}_j$. We update the emission distribution parameters by computing their sufficient statistics. For discrete distributions, the sufficient statistics are the sum of weights $p_j^{(t)}$ over non-missing samples, the sum of weights over all samples, and the weighted sum of each feature over non-missing samples. For continuous distributions, we must also compute the weighted sum of the square of each feature over non-missing samples. The updates are straightforward; we omit them due to space constraints and provide our implementation online.

Violation of model assumptions. Model fitting will converge regardless of whether there are actually cycles in the data, although convergence may be slow if model assumptions are seriously violated. Fit should be assessed by comparing model parameters and the outputs discussed in Section 4.3 with prior knowledge about the cycle to determine the plausibility of the fitted model.

Parameter initialization. CyHMMs are initialized with a hypothesized cycle length. In our simulations (Section 5), we found that CyHMMs are fairly robust to inaccurate specification of this parameter: they converged to the true length as long as they were initialized within roughly 50% of the true length. In real-world datasets, this is a reasonable constraint: the true cycle length is often at least approximately known (as it is in the two real-world case studies we consider in Section 6). To further increase robustness, the model can be fit using a range of cycle length initializations, and the model with the best log likelihood can then be used; we confirm on simulated data that this works reliably.

Scalability. Three features allow our implementation to scale to real-world datasets: we specifically adapt our implementation to use an optimized HMM library [54]; model fitting can be easily parallelized because the forward-backward computations take up most of the computation time and are independent for each time series; and we find that in real-world applications a relatively small number of hidden states suffices to capture realistic dynamics. Consequently, as we describe in Section 6, we are able to fit a CyHMM model in 3 minutes (using a computer with 16 cores) on an activity dataset 8 million total observations; we are able to fit a menstrual cycles dataset with 22 million total observations in 18 minutes (Figure 2). (By "observation", we here mean a reading of one feature for one individual at one timestep). Both these datasets were provided by companies with large user bases and we needed to perform no filtering for scalability; all filtering was performed only to select active users who could provide rich data. For extremely large datasets, a random sample of users could also be used to fit the model, since statistical power would not be a concern.

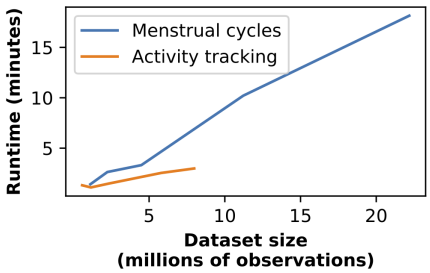

Figure 2: Time to fit CyHMM models on our real-world datasets. To provide runtimes on a range of sample sizes, we take random subsamples of the dataset. The menstrual cycles dataset is larger, so we can extrapolate further.

4.3 Inference of cycle properties

Cycle length. To infer cycle length for each individual, we fit a CyHMM to the entire population, then use the fitted model to infer the Viterbi path—that is, the most likely sequence of hidden states—for each individual from their observed data. We define the inferred cycle length as the time to return to the same hidden state in the Viterbi path.

Feature trajectories. We use CyHMMs to infer how features fluctuate (increase or decrease) over the course of a cycle as follows. After fitting the CyHMM, we assume all users are in a single substate and propagate the state vector forward by multiplying by the learned transition matrix. We compute the expected value of each feature at each timestep using the learned emission model.

Feature variabilities. We define feature variability to be how much a feature fluctuates over the course of a cycle and compute it as follows. Given the inferred trajectory for feature k—a value V_{tk} at each timestep t less than or equal to the cycle length L—we define the feature variability Δ_k as how much the feature varies relative to its mean μ_k: $\Delta_k = \frac{1}{L}\sum_{t=1}^{L}\left|\frac{V_{tk}-\mu_k}{\mu_k}\right|$. (We divide by the mean to ensure features with different scales can be compared; this definition of Δ_k is similar to the coefficient of variation [10].) Our definition of feature variability is not equivalent to merely computing the standard deviation of all observations of a feature: for example, if our feature is "person carries umbrella", its variability over the menstrual cycle will be zero because rain (and umbrellas) do not vary with hormone state.

Clustering. Previous work has found that clustering time series can identify important heterogeneity [38]; we therefore extend our model to divide individuals into clusters with similar cycle progression patterns. After initializing the clustering by dividing individuals into C clusters[3], we use an EM procedure with hard cluster assignment, alternating between two steps until convergence: 1) for each cluster c, fit a separate CyHMM $\mathcal{M}_c$ using only individuals in that cluster; 2) reassign each individual i to the cluster c whose model maximizes the log likelihood of $X^{(i)}$.

5 EVALUATION ON SIMULATED DATA

We first validate the performance of CyHMMs on synthetic data before examining their application to two real-world datasets.

5.1 Synthetic data generation

We simulate a population where the data for each individual consists of a multivariate time series with missing data. We briefly describe the simulation procedure here and make the full implementation available online. To test our model's robustness to model misspecification, we do not generate data using the CyHMM generative model. In our simulation, each individual progresses through a cycle, and both the values for each feature and the probability that data for the feature is missing vary *sinusoidally* as a function of where the individual is in their cycle. We examine the effects of altering the maximum individual time series length ($T_{\max}$), the amount of noise in the data (σ_n), the probability data is missing (p_{missing}), and the between-user and within-user variation in cycle length (σ_b and σ_w). We generate simulated data in two stages. First, we simulate the fraction of the way each person is through their cycle at each timestep, which we refer to as "cycle position"; cycle position can be thought of as a continuously varying cycle state. Second, we generate observed data given the person's cycle position.

Simulating cycle position. We draw each individual's average cycle length from $L^{(i)} \sim N(L, \sigma_b^2)$, where L is the average population cycle length and σ_b controls *between individual* variation in cycle length; we set $L = 30$ days in our simulations. We draw each individual's starting cycle day from a uniform distribution over possible cycle days; each subsequent day, their cycle day advances by 1, unless they have reached their current cycle length, in which case they go back to 0 and we draw the length for their next cycle from $N(L^{(i)}, \sigma_w^2)$. σ_w controls *within individual* variation in cycle length. We define an individual's *cycle position* $\phi_t^{(i)} \in [0, 1]$ to be the fraction of the way the individual is through their current cycle: *i.e.*, cycle day divided by current cycle length.

Simulating time series given cycle position. Given individual i's cycle position at timepoint t, $\phi_t^{(i)}$, we generate the values for a feature k, $X_{tk}^{(i)}$, as follows. We generate both binary and continuous data to evaluate our model under a wide variety of conditions. For continuous features, we allow the probability each feature is observed at each timestep, $p_{tk}^{(i)}$, to vary sinusoidally in $\phi_t^{(i)}$. We draw whether the feature is observed from Bernoulli($p_{tk}^{(i)}$). If the feature is observed, we set its value to a second sinusoidal function of $\phi_t^{(i)}$. For both sinusoids, we draw individual-specific coefficients to allow for variation in cycle dynamics across individuals. Our emission model for binary features is similar to the continuous model except that at each timestep, a single Bernoulli draw determines whether *all* features are missing, and we draw each feature from a Bernoulli whose emission probability varies sinusoidally in cycle position. Our full procedure for generating simulated data is available online.

5.2 Inference of cycle length

A basic question about any cycle is its length, both in individuals and in the population as a whole. We thus first evaluate how well CyHMMs infer the true cycle length for each individual as compared to baselines. Our evaluation metric is mean error.

Baselines. Based on the previous literature we describe in Section 2, we compare to baselines from all three major areas of related work: classical techniques (Fourier transform and autocorrelation), data-mining techniques (Partial Periodicity Detection) and circadian rhythm detection (the ARSER algorithm):

- **Discrete Fourier transform** [9]: for each feature, we perform a discrete Fourier transform and take the period of the peak with the largest amplitude. We take the median of these periods across all features. (Fourier transforms are generally used on univariate data; by taking the median, we combine data from all features.)

[3] To initialize the clustering, we choose a random subset of $\mathcal{S}$ individuals and fit one CyHMM $\mathcal{M}_s$ for each individual. We then loop over all individuals in the dataset and compute the log likelihood $\mathcal{L}_s^{(i)}$ for each individual i under each model $\mathcal{M}_s$. We run k-means on the matrix of these likelihoods to initially divide individuals into C clusters (z-scoring each individual to control for the fact that, for example, individual time series vary in length and thus log likelihoods may not be directly comparable). This initialization procedure is similar to [56] but their procedure is quadratic in the number of individuals, rendering it too slow for large real-world datasets (Section 6).

- **Autocorrelation** [51]: for each feature, we compute the lag which produces the maximum autocorrelation in that feature. We take the median of these lags across all features.
- **Partial Periodicity Detection** [40] (binary data only): we apply Ma and Hellerstein's χ^2 test for partial periodicity detection[4] to each individual feature. The algorithm returns a list of statistically significant periods; we take the median of these periods. The algorithm relies on specification of a δ parameter, which controls the time tolerance in period length; we evaluate $\delta = 2, 5, 10$, find the algorithm yields the most accurate cycle length inference with $\delta = 2$, and report results for this parameter setting.
- **ARSER circadian rhythm detection** [63]: the ARSER algorithm fits a sinusoidal model to each individual feature[5], and reports a p-value for periodicity. Consistent with the original authors, we filter for features which show statistically significant periodicities after multiple hypothesis correction; we then take the period with the largest amplitude for each feature and aggregate across multiple features by taking the median. We find that setting a statistical significance threshold of $p = .01$ yields the most accurate cycle length inference, and report results for this parameter setting. We note that the original implementation does not scale to datasets of our size, so to evaluate it we must make two modifications: we parallelize it to use multiple cores, and we fit the autoregression models using only the Yule-Walker and Burg methods, because the MLE method does not scale.

In assessment, we filter out periods which are shorter than 5 days or longer than 50 days (the true cycle length is 30 days). This increases accuracy by, for example, preventing the autocorrelation baseline from returning implausibly short lags (which in real-world settings an analyst would filter out) which degrade performance. We run 100 binary simulation trials and 100 continuous simulation trials, systematically varying the maximum individual time series length ($T_{\max}$) from 90 to 180 days, the amount of noise in the data (σ_n) from 5 to 50, the probability data is missing (p_{missing}) from 0% to 90%, and the between-user and within-user variation in cycle length (σ_b and σ_w) from 1 to 10 days.

Results. CyHMMs have lower mean error than do baselines on both binary and continuous data (Table 1). Across all 200 trials, they reduce the mean error in cycle length inference over the baseline with lowest mean error, autocorrelation, by 58%. Their inferred cycle lengths for each person are also better correlated with the true cycle lengths for each person ($r = .57$) than the baseline with the best correlation (ARSER, $r = .19$) demonstrating that they can better identify individual heterogeneity.

When we examine the effects of altering specific simulation parameters, we find that CyHMM performance, as measured by correlation with true cycle length, improves as noise decreases, fraction of missing data decreases, maximum time series length increases, variability in cycle length *between* individuals increases, and variability in cycle length *within* individuals decreases. We also find that the superior performance of CyHMMs over baselines occurs in part because CyHMMs are more robust to noise and missing data. In very low noise settings, several baselines perform comparably to CyHMMs (autocorrelation, ARSER, and Fourier), but baseline performance rapidly degrades in more realistic scenarios as noise increases. It is logical that CyHMMs would be more robust to noise, because they share information across individuals and across features, while baselines do not. Similarly, CyHMMs should be more robust to missing data because they explicitly incorporate it into the generative model and use it to infer cycle state. We also note that it is striking that CyHMMs can outperform ARSER and Fourier transforms even though both are based on sinusoidal models, and our simulated data is drawn from sinusoids.

[4]https://github.com/jcborges/PeriodicEventMining
[5]https://github.com/cauyrd/ARSER

	All trials	Binary	Continuous
CyHMM	**2.5 days**	**3.1 days**	**1.9 days**
Autocorrelation	5.9 (58%)	6.8 (55%)	4.9 (62%)
ARSER	10.5 (76%)	13.7 (78%)	7.2 (74%)
Fourier	13.3 (81%)	16.0 (81%)	10.6 (83%)
Partial Periodicity Detection	Binary only	19.4 (84%)	Binary only

Table 1: Mean error in inferring cycle length across simulation trials. CyHMMs have lower mean error than all baselines. Relative error improvement of CyHMM over baselines shown in parentheses. We order methods by their mean error. Partial Periodicity Detection can only be applied to binary data.

5.3 Inference of feature trajectories

We compute feature trajectories and variabilities as described in Section 4.3, and assess how well the variabilities computed from the inferred trajectories correlate with the variabilities computed from the true trajectories. (To compute the true trajectories, we compute the mean value of each symptom on each cycle day.) The mean correlation between the true and inferred variabilities is 0.97 across all binary simulations and 0.98 across all continuous simulations, demonstrating that CyHMMs are able to correctly infer feature variabilities. Because the baselines discussed in Section 5.2 are not generative models designed to infer feature trajectories, we do not compare to them.

6 EVALUATION ON REAL-WORLD DATA

We demonstrate the applicability of CyHMMs by using them to infer cycle characteristics in two real-world health datasets: a menstrual cycles dataset, and an exercise and activity tracking dataset. We choose these datasets for two reasons: first, they have information on true cycle state, allowing us to compare our model to existing baselines, and second, they track important health-related features[6].

6.1 Datasets

Menstrual cycles dataset. The human menstrual cycle is fundamental to women's health; features of the cycle have been linked to cancer [23], depression [19], and sports injuries [59], and the cycle has been proposed as a vital sign [4]. We obtained data from one of the most popular menstrual cycle tracking apps. Upon logging into the app, users can record *binary* features including period/cycle start (*e.g.*, light, moderate, or heavy bleeding), mood (*e.g.*, happy, sad, or sensitive), pain (*e.g.*, cramps or headache) or sexual activity (*e.g.*, protected or unprotected sex). We define the start of each menstrual cycle as the start of period bleeding. We confirmed that

[6]Data from both companies has been used in previous studies and users are informed that their data may be used. Because all data analyzed is preexisting and de-identified, the analysis is exempt from IRB review.

statistics of the dataset were consistent with existing literature on the menstrual cycle: the average cycle duration was 28 days, consistent with previous investigations, and features that peaked before the cycle start were consistent with previous investigations of premenstrual syndrome [50, 64].

This dataset has a number of traits which make it an useful test case for cycle inference methods. It has a rich set of features for each user, a large number of users, and variation in cycles both between and within users: menstrual cycle length varies from person to person, and even from cycle to cycle, as do cycle symptoms. Most importantly, it contains ground truth on the true cycle start times for each user (as indicated by bleeding). While CyHMMs are more generally applicable to cases when the true cycle starts are not known, true cycle starts are essential for validating that the method works and comparing it to other methods.

We fit a CyHMM using features shown by the tracking app's default five categories—mood, sleep, sex, pain, and energy—because many of the other features had very sparse data. Importantly, we did not provide the model with any features directly related to cycle start, like bleeding, so the model was given no information about when cycles started. In total this left us with 19 binary features for each day. We filtered for users that logged a feature using the app at least 20 times and had a timespan of at least 50 days where they used the app regularly (logging at least once every two days). After filtering, our analysis included 22 million observations (*i.e.*, measurements of one feature for one user at one timestep) from 9,885 app users. We fit a CyHMM with four hidden states because the menstrual cycle is often divided into four latent hormonal phases [7]. (To study the robustness of our model to this parameter, we also examined results using three or five states.) If a user logged no features on a given day, we considered data for that day to be unobserved because, as explained above, we did not know whether the user truly had no symptoms or merely neglected to log them.

Activity tracking dataset. We obtained data from a large activity tracking mobile app that has been previously used in studies of human health [2, 3, 55]. For each user on each day, our dataset included sleep start time, sleep end time, steps taken, and total calories burned. Each feature had missing data caused, for example, by a user forgetting to log their sleep. Our analysis included all regular app users that logged into the app to record a measurement at least 20 times, had a timespan of at least 50 days where they used the app regularly (logging at least once every two days) and had less than 50% missing data in every feature; in total, after filtering, our data consisted of 8 million observations from 6,882 users. Because we are interested in cycles, we removed long-term time trends (*e.g.*, people using the app often lose weight) by subtracting off the centered moving average using a two-week window around each day. This is analogous to the trend-cycle decomposition often used in time-series analysis [20].

Previous research has found that people's activity patterns show weekly cycles: for example, they sleep and wake later on the weekends [43]. We thus assessed how well CyHMMs could model this weekend effect from time series data without being provided with the day of the week: essentially, testing whether CyHMMs could recover weekly cycle patterns without knowing the day of the week. We fit a Gaussian CyHMM with two states to correspond to weekday and weekend using sleep start time, sleep end time, steps taken, and calories burned as continuous features. We verified that these features did in fact show weekly cycles in the data: individuals go to bed and wake up later on average on the weekends, with steps taken and calories burned remaining fairly constant across weekends and weekdays.

6.2 Analysis of menstrual cycles

We used ground truth data to evaluate how well CyHMMs could recover basic features of the menstrual cycle: cycle length, trajectory of each feature over the course of the cycle, most variable features, and clusters of individuals with different symptom patterns.

Method	Population cycle length	Median error	Mean error
Ground truth	28.0 days	0.0 days	0.0 days
CyHMM	**29.0**	**2.0**	**3.6**
Autocorrelation	19.0	9.0	9.9
Partial Periodicity Detection	21.9	8.4	10.0
Fourier transform	14.8	12.7	11.9
ARSER	20.9	13.3	12.7

Table 2: Comparison of our model to baselines on menstrual cycle data for users who recorded five or more cycles. Errors reported are averaged across all users; the error for each user is the difference between the true cycle length for each user and the inferred cycle length for that user. We order methods by their mean error.

6.2.1 Inference of cycle length. As with simulated data, we assessed how well CyHMMs could infer the true cycle length for each user on the menstrual cycles dataset, defining the true cycle length for each user to be their average gap between period starts. To ensure we had reliable data for each user, we assessed performance on users who recorded at least 5 cycles. We compared our model to the baselines described in Section 5.2. CyHMMs inferred the cycle length for each user more accurately than did the baselines (Table 2) and were also closer to the population average cycle length[7]. CyHMMs had a mean error that was 63% lower, and a median error that was 78% lower, than autocorrelation, the baseline with the lowest mean error.

Inference of feature trajectories. We inferred feature trajectories using the methodology described in Section 4.3. In Figure 3 we visualize the inferred feature trajectories for pain, emotion, and sleep features. To provide context for the trajectories, we also plot the true cycle start day (when period bleeding starts), although this information is not given to the model. The model correctly infers that mood and pain features increase just before the cycle starts, consistent with existing literature on premenstrual syndrome [50, 64]. Finally, the model infers that users are more likely to log in to record features when their cycle is starting; this is logical because many people use the app to track cycle starts, and speaks to the importance of modeling missing data.

[7] We report results using the 4-state CyHMM, but also found that 3 and 5 state CyHMMs significantly outperformed baselines. We also compared to the baseline of a CyHMM with a geometric duration distribution (rather than a Poisson duration distribution). The errors in inferred cycle length using a geometric rather than a Poisson distribution were 10%, 47%, and 129% higher for the 3, 4, and 5 state models, respectively, confirming that the Poisson distribution provided a better fit to the data.

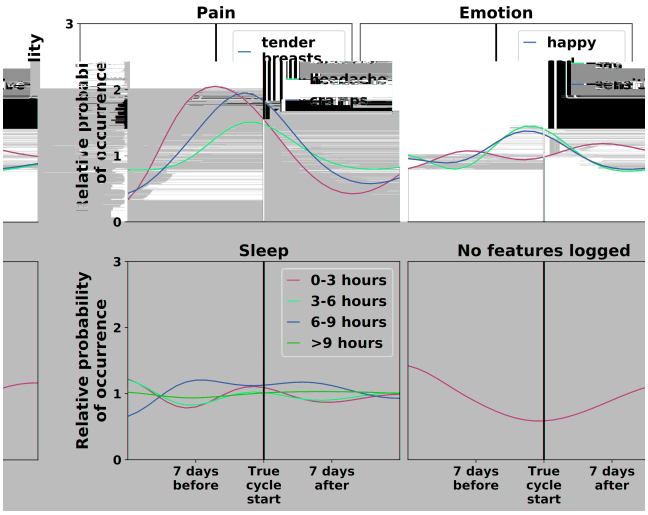

Figure 3: Model-inferred feature trajectories for a subset of menstrual cycle features. The vertical axis is the relative probability of emitting a feature (conditional on non-missing data); the horizontal axis is cycle day. For context, the vertical black line shows the day of the true cycle start, although this information is not given to the model. We center cycle start for ease of visualization. Pain and negative mood features increase just prior to cycle start; sleep features do not show consistent patterns. The relative probability that *no* features are logged ("No features logged") decreases near cycle start, because many people use the app to track their cycle starts.

Cluster	1	2	3	4	5
Emotion:happy (% of time)	55%	15%	**74%**	24%	66%
Emotion:sad	**22%**	9%	11%	9%	17%
Emotion:sensitive	**46%**	15%	27%	20%	32%
High energy	36%	7%	**53%**	16%	4%
Low energy	**43%**	11%	33%	23%	4%
Pain:cramps	**20%**	**20%**	12%	12%	15%
Pain:headache	**33%**	14%	15%	14%	17%
Pain:tender breasts	**18%**	11%	9%	9%	12%
High sex drive	**20%**	7%	6%	6%	11%
Had protected sex	5%	**15%**	4%	5%	5%
Had unprotected sex	10%	**25%**	4%	8%	6%
Had withdrawal sex	**16%**	8%	2%	3%	7%
Slept 0-3 hours	**5%**	1%	1%	1%	0%
Slept 3-6 hours	**40%**	8%	18%	21%	2%
Slept 6-9 hours	39%	13%	**63%**	52%	4%
Slept 9+ hours	10%	3%	**13%**	9%	1%
Features logged per non-missing day	**4.6**	1.9	3.7	2.5	2.1
Fraction of days with no data	5%	**68%**	4%	18%	10%
Recorded doctor's appt	**37%**	15%	29%	27%	15%
Recorded ob/gyn appt	**24%**	14%	17%	18%	11%
Took pain medication	**36%**	17%	27%	27%	16%
Logged birth control pill	38%	**85%**	29%	49%	33%
Negative pregnancy test	15%	13%	9%	13%	**16%**
Positive pregnancy test	9%	5%	6%	6%	**10%**
Age	21.9	21.7	20.6	**23.9**	20.8
Cycle length, days	28.2	29.0	28.8	28.3	**29.4**

Table 3: Clustering of menstrual cycle data. Only the top set of user characteristics (above the double line) is accessible by the model, but the clustering also correlates with differences in medical history, age, and true cycle length (bottom set of characteristics) which are not provided to the model. All differences between clusters are statistically significant ($p < 0.001$, categorical F-test). The largest value in each row is shown in bold.

Feature variability. We also found that CyHMMs correctly recovered which features showed the most variability over the course of the cycle. We compare the variabilities computed from the true feature trajectories to the variabilities computed from the inferred feature trajectories as described in Section 4.3. (We computed the true feature trajectories by aligning all users to their period starts, and computing the fraction of users experiencing a given feature on each day relative to period start). The true variabilities were highly correlated with the inferred variabilities ($r = 0.86$): *e.g.*, the model correctly recovered the fact that pain features showed greater variability than sleep features.

Clustering of similar individuals. We divided individuals into five clusters with similar feature patterns using the methodology described in Section 4.3. We chose the number of clusters using the elbow method [57]: beyond five clusters, both the train and test log likelihoods increased much more slowly. Our clustering correlates with individual-specific features not given to the model (Table 3; all features below the double horizontal line), like medical history and age. We report only features which show statistically significant differences (categorical F-test $p < 0.001$), revealing distinct groups:

- **Cluster 1: Severe symptom users.** This cluster (Table 3, first column), is most likely to report going to the doctor or gynecologist or taking pain medication (rows below double horizontal line). Our model is able to identify this subcluster even though it is given no information on doctors' visits. These users also have the highest rate of pain features, are more likely to report negative emotions and less sleep, and are most likely to report having a high sex drive (rows above double horizontal line).
- **Cluster 2: Birth control users.** These users primarily use the app to keep track of whether they have taken birth control ("Logged birth control pill" row)—again, not information given to the model. Hence, they are less likely to log other features and have a high fraction of missing data ("Fraction of days with no data" row), but likely to be on birth control. Consistent with this, they have a higher rate of having unprotected sex than do other users, and the lowest rate of positive pregnancy tests.
- **Cluster 3: Happy users.** This group of users is most likely to report positive features like happy emotion (top row), high energy, and sleeping 6-9 hours.
- **Cluster 4: Low-emotion users.** This group of users is relatively unlikely to log any emotion features (top three rows).
- **Cluster 5: Infrequent loggers.** This cluster of users appears to put less effort into using the app; when they record features, they record the fewest of any group besides the birth control group ("Features logged per non-missing day" row).

The fact that the clustering correlates with features not used in the clustering—like birth control, doctors' appointments, and age—indicates that CyHMMs are correctly identifying true population heterogeneity. For example, the model is able to identify "users more likely to be on birth control" (information it is not given) as "users with a high probability of missing data" (something it

explicitly models). Recovering the birth control subpopulation is essential for accurately modeling the menstrual cycle. Previous work has found that women on birth control experience different menstrual cycle symptom progression—for example, less variability in mood over the course of the cycle [45, 48]. (We note that the fact that women in the birth control cluster are relatively unlikely to bother logging emotion symptoms is consistent with this lack of variability.) Recovering the birth control subpopulation is also important for designing health interventions. For example, women on birth control require different pharmacologic treatment [30] and, because their hormone cycles are different, have different risks of hormone-related cancers [44, 46].

6.3 Analysis of cycles in activity tracking

Inference of population cycle length. We define the model-inferred cycle lengths as in Section 4.3. Since there are no ground truth cycle lengths for all individuals in this dataset—unlike with the menstrual cycles dataset, there is no clear physiological feature with which to define cycle starts—a supervised comparison to baselines is impossible in this setting. Instead, we directly evaluate the plausibility of CyHMM cycle length inference. When we apply CyHMMs to our dataset, the model infers that the most frequent cycle length in the population as a whole is 7 days (Figure 4). Inspecting the model, the two learned hidden states also clearly correspond to weekday and weekend, even though day of the week is not provided to the model: the "weekend hidden state" has an inferred average duration of 2.2 days (as compared to the true weekend length of 2 days), while the "weekday hidden state" has an inferred average duration of 4.9 days (as compared to the true weekday length of 5 days). The weekend hidden state has a later inferred mean for sleep start and sleep end than the weekday hidden state, consistent with the fact that individuals sleep later on the weekends. On timesteps when the model infers an individual is in the weekend state, it is twice as likely to be the true weekend. All these facts indicate that the model is successfully recovering weekday/weekend cycles. This is consistent with previous findings that human activity shows weekly cycles [43].

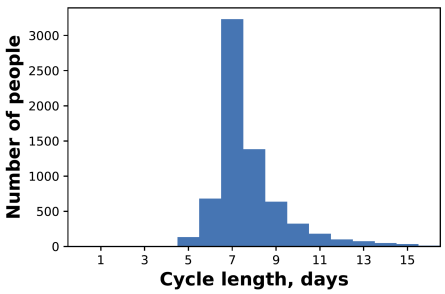

Figure 4: Histogram of cycle lengths for individuals in the activity tracking dataset. The cycle length for each individual is the modal cycle length for that individual.

Clustering of similar individuals. We next investigated whether CyHMMs could cluster individuals into groups that correlated with fundamental features like BMI, age, and gender without being provided with these features to use in the clustering. Our clustering shows statistically significant correlations with all these features ($p < 0.001$, categorical F-test). For example, we identify one cluster with the highest proportion of males (78%) with the highest BMIs (26.9); another cluster, essentially its opposite, has the lowest proportion of males (48%) and the lowest BMIs (25.4). (Cluster differences between BMIs remain significant when adjusting for age and sex via linear regression.) None of these features are given to the CyHMM in performing the clustering, indicating that the model is successfully identifying true population heterogeneity. Our finding is consistent with previous work that finds that weekend-weekday sleep differences correlate with important health metrics like body mass index (BMI) [53]. CyHMMs can thus identify subpopulations that differ along fundamental health metrics like BMI, potentially allowing for unsupervised discovery of behavioral risk factors to inform monitoring of obesity.

7 CONCLUSION

Modeling cycles in time series data is important for understanding human health, but it is challenging because cycle starts are rarely labeled. Here we present CyHMMs, which take as input a multivariate time series for each individual in a population, infer what latent cycle state an individual is in at each timestep, and use this inferred state to recover fundamental cycle characteristics. Evaluating our method on both simulated data and two real-world datasets with ground-truth information on cycle states, we find CyHMMs infer the true cycle length more accurately than do baselines. CyHMMs can also infer cycle characteristics that baselines cannot: we show they can infer how each observed feature changes over the course of the cycle, accurately recover the most variable features, and find clusters of individuals with distinct cycle patterns. While we evaluated CyHMMs on datasets in which cycle starts are known in order to validate our method, they are designed to be applied to datasets on which cycle starts are unknown: for example, sentiment in individual Twitter feeds.

CyHMMs assume that observed data is generated by a single latent state that progresses cyclically. Future work could attempt to relax this assumption while retaining the interpretability it provides. Natural extensions to the CyHMM model might retain its core idea of a cyclic latent state while increasing the model's expressive power: for example, by using a multidimensional hidden state as factorial HMMs [24] or LSTMs [31] do, with some cyclic and some unconstrained dimensions. A multidimensional hidden state could capture data that includes multiple cycles, or both a cycle and a time trend. These extensions flow from the fact that CyHMMs, by using a generative model with hidden state, provide a natural and flexible way to model cycles fundamental to human health.

Code availability. https://github.com/epierson9/cyclic_HMMs.

Acknowledgments. We thank David Hallac, Chris Olah, Nat Roth, Marinka Zitnik, the apps that provided data, and the reviewers for their valuable feedback. E.P. was supported by Hertz and NDSEG Fellowships. T.A. was supported by National Institutes of Health (NIH) grant U54 EB020405 and a SAP Stanford Graduate Fellowship.

REFERENCES

[1] T. Althoff, E. Horvitz, R. W. White, and J. Zeitzer. Harnessing the web for population-scale physiological sensing: A case study of sleep and performance. *WWW*, 2017.

[2] T. Althoff, P. Jindal, and J. Leskovec. Online actions with offline impact: How online social networks influence online and offline user behavior. In *WSDM*, 2017.

[3] T. Althoff, R. Sosic, J. L. Hicks, A. C. King, S. L. Delp, and J. Leskovec. Large-scale physical activity data reveal worldwide activity inequality. *Nature*, 2017.

[4] American Academy of Pediatrics, American College of Obstetricians and Gynecologists, et al. Menstruation in girls and adolescents: using the menstrual cycle as a vital sign. *Pediatrics*, 2006.

[5] W. G. Aref, M. G. Elfeky, and A. K. Elmagarmid. Incremental, online, and merge mining of partial periodic patterns in time-series databases. *TKDE*, 2004.

[6] C. Berberidis, I. Vlahavas, W. G. Aref, M. Atallah, and A. K. Elmagarmid. On the discovery of weak periodicities in large time series. In *PKDD*, 2002.

[7] S. L. Berga and S. Yen. Circadian pattern of plasma melatonin concentrations during four phases of the human menstrual cycle. *Neuroendocrinology*, 1990.

[8] J. A. Bilmes et al. A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden Markov models. *International Computer Science Institute*, 1998.

[9] R. N. Bracewell and R. N. Bracewell. *The Fourier transform and its applications*. 1986.

[10] C. E. Brown. Coefficient of variation. In *Applied multivariate statistics in geohydrology and related sciences*. 1998.

[11] G. Celeux and J.-B. Durand. Selecting hidden Markov model state number with cross-validated likelihood. *Computational Statistics*, 2008.

[12] A. K. Chanda, S. Saha, M. A. Nishi, M. Samiullah, and C. F. Ahmed. An efficient approach to mine flexible periodic patterns in time series databases. *Engineering Applications of Artificial Intelligence*, 2015.

[13] P. Chaovalit, A. Gangopadhyay, G. Karabatis, and Z. Chen. Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys (CSUR)*, 2011.

[14] L. Chiazze, F. T. Brayer, J. J. Macisco, M. P. Parker, and B. J. Duffy. The length and variability of the human menstrual cycle. *JAMA*, 1968.

[15] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *Transactions on Information Theory*, 1990.

[16] A. Deckard, R. C. Anafi, J. B. Hogenesch, S. B. Haase, and J. Harer. Design and analysis of large-scale biological rhythm studies: a comparison of algorithms for detecting periodic signals in biological data. *Bioinformatics*, 2013.

[17] M. Dewar, C. Wiggins, and F. Wood. Inference in hidden Markov models with explicit state duration distributions. *IEEE Signal Process. Lett.*, 2012.

[18] D. F. Dinges. An overview of sleepiness and accidents. *Journal of Sleep Research*, 1995.

[19] J. Endicott. The menstrual cycle and mood disorders. *Journal of Affective Disorders*, 1993.

[20] W. Fellner. Trends and cycles in economic activity. 1956.

[21] J. Ferguson. Variable duration models for speech. In *Proc. Symp. on the Application of Hidden Markov Models to Text and Speech*, 1980.

[22] R. L. Findling, B. L. Gracious, N. K. McNamara, E. A. Youngstrom, C. A. Demeter, L. A. Branicky, and J. R. Calabrese. Rapid, continuous cycling and psychiatric co-morbidity in pediatric bipolar I disorder. *Bipolar disorders*, 2001.

[23] M. Garland et al. Menstrual cycle characteristics and history of ovulatory infertility in relation to breast cancer risk in a large cohort of US women. *American Journal of Epidemiology*, 1998.

[24] Z. Ghahramani and M. I. Jordan. Factorial hidden Markov models. In *NIPS*, 1996.

[25] C. Giannella, J. Han, J. Pei, X. Yan, and P. S. Yu. Mining frequent patterns in data streams at multiple time granularities. *Next Generation Data Mining*, 2003.

[26] S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 2011.

[27] J. Han, G. Dong, and Y. Yin. Efficient mining of partial periodic patterns in time series database. *ICDE*, 1999.

[28] J. Han, W. Gong, and Y. Yin. Mining segment-wise periodic patterns in time-related databases. In *KDD*, 1998.

[29] J. Han, J. Pei, Y. Yin, and R. Mao. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *DMKD*, 2004.

[30] T. Hassan. Pharmacologic considerations for patients taking oral contraceptives. *Connecticut Dental Student Journal*, 1987.

[31] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 1997.

[32] Y.-H. Hu, C.-F. Tsai, C.-T. Tai, and I.-C. Chiang. A novel approach for mining cyclically repeated patterns with multiple minimum supports. *Applied Soft Computing*, 2015.

[33] M. E. Hughes, J. B. Hogenesch, and K. Kornacker. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of Biological Rhythms*, 2010.

[34] R. U. Kiran, H. Shang, M. Toyoda, and M. Kitsuregawa. Discovering recurring patterns in time series. In *EDBT*, 2015.

[35] O. Kostakis, N. Tatti, and A. Gionis. Discovering recurring activity in temporal networks. *DMKD*, 2017.

[36] E. A. Lenton, B. Landgren, and L. Sexton. Normal variation in the length of the luteal phase of the menstrual cycle: identification of the short luteal phase. *British Journal of Obstetrics and Gynecology*, 1984.

[37] S. E. Levinson. Continuously variable duration hidden Markov models for automatic speech recognition. *Computer Speech & Language*, 1986.

[38] L. Li and B. A. Prakash. Time series clustering: Complex is simpler! In *ICML*, 2011.

[39] Z. Li, J. Wang, and J. Han. ePeriodicity: Mining event periodicity from incomplete observations. *TKDE*, 2015.

[40] S. Ma and J. L. Hellerstein. Mining partially periodic event patterns with unknown periods. In *International Conference on Data Engineering Proceedings*, 2001.

[41] B. Mc Mahon et al. Seasonal difference in brain serotonin transporter binding predicts symptom severity in patients with seasonal affective disorder. *Brain*, 2016.

[42] J. M. Mitchison. *The Biology of the Cell Cycle*. 1971.

[43] T. H. Monk, D. J. Buysse, L. R. Rose, J. A. Hall, and D. J. Kupfer. The sleep of healthy people – a diary study. *Chronobiol. Int*, 2000.

[44] S. A. Narod, H. Risch, R. Moslehi, A. Dørum, S. Neuhausen, H. Olsson, D. Provencher, P. Radice, G. Evans, S. Bishop, et al. Oral contraceptives and the risk of hereditary ovarian cancer. *New England Journal of Medicine*, 1998.

[45] K. A. Oinonen and D. Mazmanian. To what extent do oral contraceptives influence mood and affect? *Journal of Affective Disorders*, 2002.

[46] C. G. on Hormonal Factors in Breast Cancer et al. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53,297 women with breast cancer and 100,239 women without breast cancer from 54 epidemiological studies. *The Lancet*, 1996.

[47] S. Orlando, P. Palmerini, R. Perego, and F. Silvestri. Adaptive and resource-aware mining of frequent sets. In *ICDM*, 2002.

[48] K. E. Paige. Effects of oral contraceptives on affective fluctuations associated with the menstrual cycle. *Psychosomatic Medicine*, 1971.

[49] T. Partonen and J. Lönnqvist. Seasonal affective disorder. *The Lancet*, 1998.

[50] T. Pearlstein, K. A. Yonkers, R. Fayyad, and J. A. Gillespie. Pretreatment pattern of symptom expression in premenstrual dysphoric disorder. *Journal of affective disorders*, 2005.

[51] R. S. Pindyck and D. L. Rubinfeld. *Econometric models and economic forecasts*. 1998.

[52] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989.

[53] T. Roenneberg, K. V. Allebrandt, M. Merrow, and C. Vetter. Social jetlag and obesity. *Current Biology*, 2012.

[54] J. Schreiber. Pomegranate: fast and flexible probabilistic modeling in Python. *arXiv preprint arXiv:1711.00137*, 2017.

[55] A. Shameli, T. Althoff, A. Saberi, and J. Leskovec. How gamification affects physical activity: Large-scale analysis of walking challenges in a mobile application. In *WWW*, 2017.

[56] P. Smyth. Clustering sequences with Hidden Markov Models. *NIPS*, 1997.

[57] R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *JRSS-B*, 2001.

[58] X. Wang, D. Sontag, and F. Wang. Unsupervised learning of disease progression models. In *KDD*, 2014.

[59] E. M. Wojtys, L. J. Huston, T. N. Lindenfeld, T. E. Hewett, and M. L. V. Greenfield. Association between the menstrual cycle and anterior cruciate ligament injuries in female athletes. *The American Journal of Sports Medicine*, 1998.

[60] G. Wu, J. Zhu, J. Yu, L. Zhou, J. Z. Huang, and Z. Zhang. Evaluation of five methods for genome-wide circadian gene identification. *Journal of Biological Rhythms*, 2014.

[61] J. Yang, W. Wang, and P. S. Yu. Infominer: mining surprising periodic patterns. In *Knowledge Discovery and Data Mining*, 2001.

[62] J. Yang, W. Wang, and P. S. Yu. Mining asynchronous periodic patterns in time series data. *TKDE*, 2003.

[63] R. Yang and Z. Su. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*, 2010.

[64] K. A. Yonkers, P. S. O'Brien, and E. Eriksson. Premenstrual syndrome. *The Lancet*, 2008.

[65] S.-Z. Yu. Hidden semi-Markov models. *Artificial Intelligence*, 2010.