Analyzing physiological signals recorded with a wearable sensor across the menstrual cycle using circular statistics

Krystal Sides[1], Grentina Kilungeja[1], Matthew Tapia[1], Patrick Kreidl[1], Benjamin H. Brinkmann[2] and Mona Nasseri[1,2]*

[1]School of Engineering, University of North Florida, Jacksonville, FL, United States, [2]Bioelectronics Neurophysiology and Engineering Laboratory, Department of Neurology, Mayo Clinic, Rochester, MN, United States

This study aims to identify the most significant features in physiological signals representing a biphasic pattern in the menstrual cycle using circular statistics which is an appropriate analytic method for the interpretation of data with a periodic nature. The results can be used empirically to determine menstrual phases. A non-uniform pattern was observed in ovulating subjects, with a significant periodicity ($p < 0.05$) in mean temperature, heart rate (HR), Inter-beat Interval (IBI), mean tonic component of Electrodermal Activity (EDA), and signal magnitude area (SMA) of the EDA phasic component in the frequency domain. In contrast, non-ovulating cycles displayed a more uniform distribution ($p > 0.05$). There was a significant difference between ovulating and non-ovulating cycles ($p < 0.05$) in temperature, IBI, and EDA but not in mean HR. Selected features were used in training an Autoregressive Integrated Moving Average (ARIMA) model, using data from at least one cycle of a subject, to predict the behavior of the signal in the last cycle. By iteratively retraining the algorithm on a per-day basis, the mean temperature, HR, IBI and EDA tonic values of the next day were predicted with root mean square error (RMSE) of 0.13 ± 0.07 (C°), 1.31 ± 0.34 (bpm), 0.016 ± 0.005 (s) and 0.17 ± 0.17 (μS), respectively.

1 Introduction

Among methods that have been used to study the menstrual cycle to detect ovulation, such as transvaginal ultrasonography, and cervical mucus inspection, the most common, and less invasive methods, are basal body temperature (BBT) tracking and/or LH-testing (Luteinizing Hormone). In normal physiology, the follicle stimulating hormone (FSH) and luteinizing hormone (LH) activate the ovary and produce follicles. The FSH motivates the growth of ovarian follicles and, of the 30–40 developing follicles, typically only one is released per month (Schmalenberger et al., 2021; Kiranmai and Lakshmi, 2021). The mature follicle then produces increasing amounts of estrogen, resulting in the LH wave (Figure 1A). Traditionally, LH-testing is used to determine if a menstrual cycle is ovulatory and when ovulation occurs. The urinary LH kit has been found to be more accurate than BBT tracking; however, a BBT chart (Figure 1B) can determine the onset of ovulation (San Roman et al.,

1995). This method is often the most popular and cost-effective for predicting ovulation, but not the most accurate and oftentimes procedurally daunting—the BBT must be measured every morning before the first urine with a thermometer (Bull et al., 2019; Coyne et al., 2000). As depicted in Figure 1A, BBT is low across the follicular phase, dips before ovulation, increases sharply at ovulation, and then remains elevated across the luteal phase. Essentially, a BBT provides information on the existence of either a biphasic or monophasic pattern, where a biphasic BBT is indicative of ovulation.

Another technique to determine the ovulation day, or fertile window, is the forward/backward-count method (Schmalenberger et al., 2021), which estimates the ovulation day as the middle of the cycle. However, this technique depends on the cycle length. In recent years, this technique has been implemented in smart-phone applications to track menstrual cycles and predict the ovulation day, or fertile window. The fertile range, however, varies broadly even with the suggested "most fertile" days and most of the publicly-available applications are inaccurate (Setton et al., 2016). Thus, tracking ovulation requires additional methodologies to improve reliability, which this study aims to do through a combinatorial

approach including analyzing physiological signals, self-reported urine test, and self-reported menses.

Few adopted methods using a combination of physiological signals have been utilized to predict ovulation (Schmalenberger et al., 2021). Signal processing techniques can be used to predict the fertile window more accurately and avoid burdensome traditional methods like LH-testing. Studies have demonstrated that signals recorded continuously with wearable sensors, such as heart rate (HR), skin temperature, and heart rate variability (HRV), respond to different phases of the menstrual cycle (Tenan et al., 2014; Alzueta et al., 2022; Schmalenberger et al., 2021; Yu et al., 2022). In particular, Schmalenberger et al. (2020) showed that HRV decreases significantly from the follicular to the luteal phase in naturally cycling subjects. Authors in (Vishrutha et al., 2012) concluded that the high frequency, or HF (0.15–0.40 Hz) component of HRV was higher in the follicular phase, while the low frequency, or LF (0.04–0.15 Hz) component, was found to be higher in the ovulatory and luteal phases.

To investigate whether physiological parameters are changing across the menstrual cycle, regression models have been used widely in literature (Shilaih et al., 2018). For example, changes in
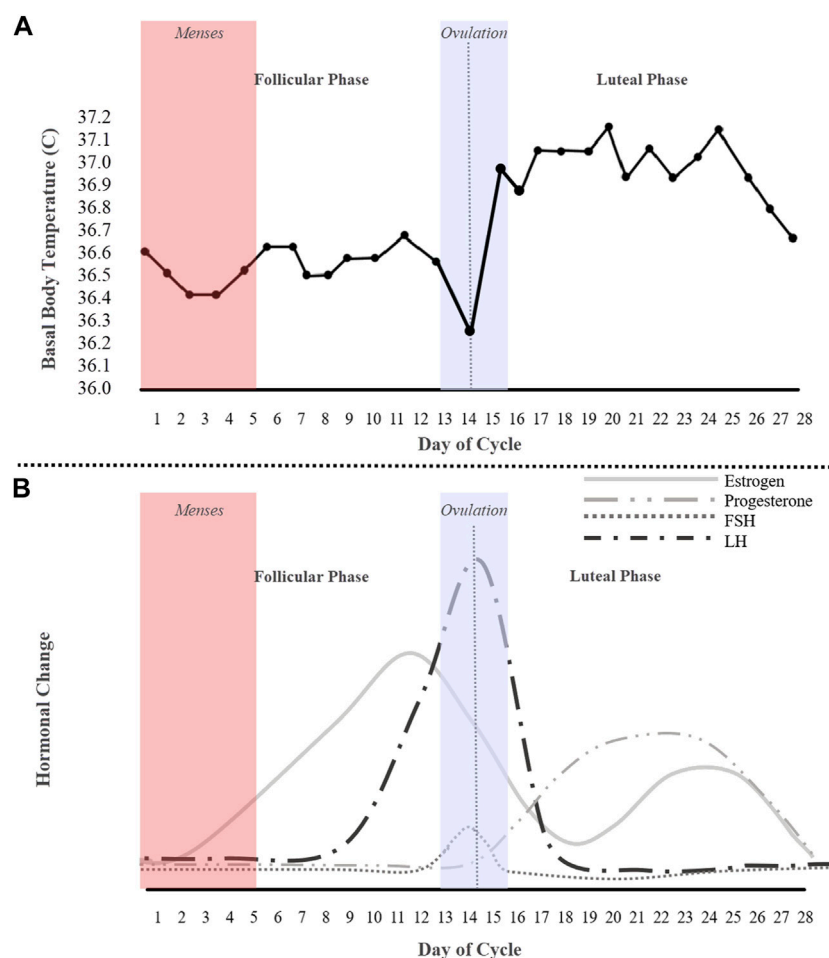
FIGURE 1
An idealized 28-day cycle illustration of **(A)** a basal body temperature (BBT) chart of an ovulating subject depicting a biphasic pattern, and **(B)** fluctuations of the ovarian hormones estrogen, progesterone, follicle stimulating hormone (FSH), and the luteinizing hormone (LH).

physiological signals such as temperature, HR and HRV recorded from the Oura ring (Ōura Health Oy, Oulu, Finland) across the four menstrual cycle phases (Menses, Ovulation, Mid-luteal and Late-luteal) were statistically tested using Hierarchical Linear Regression models in (Alzueta et al., 2022). It was concluded that there were phase-based shifts in nightly skin temperature and HR across the four phases. Another study (Goodale et al., 2019) considering a series of multilevel models with random slopes and random intercepts, demonstrated that wearable technology (Ava bracelet, Ava, Zurich, Switzerland) can detect concurrent phase-based shifts in wrist temperature, HR and respiratory rate (p < 0.001). The authors also developed an ensemble tree-based machine learning method to separate three main classes: follicular phase, fertile window and luteal phase. This resulted in an overall F-score of 0.78 and it detected a 6-day fertile window in cycles with a 90% accuracy. The classifier was trained on 11 features extracted from the physiological signals, including HR, breathing rate, wrist skin temperature and HRV. The training and testing data split was performed randomly with a 75:25 ratio.

Although many research studies focus solely on subjects with "regular" cycles (with a length of 25–35 days), Yu et al. (2022) has shown that BBT and HR were significantly higher during fertile phase than the follicular phase and peaked in the luteal phase (p < 0.001) in both regular and irregular cycles (< 25, > 35 days). In Yu et al. (2022) linear mixed models were used to assess changes in physiological signals and a probability function estimation model was developed to predict the fertile window and menses.

In this study, to examine the changes in physiological features during menstrual cycles, circular statistics were exploited. Circular statistics, known as directional statistics, are generally utilized in applications using data in the R^2 plane, not a linear scale (R) (Pewsey and García-Portugués, 2021; Landler et al., 2019), considering the periodic nature of angular measurements, such as angles, directions, or times. Circular statistics were developed first in the late 19th and 20th centuries. However, they gained more attention in the mid-20th century (Fisher, 1922), and during the latter half of the 20th century, circular statistics found applications in various fields, including biology, once it was realized that circular data was prevalent in many natural phenomena and human activities such as biological rhythms, animal migration patterns, and wind directions.

Recently, circular statistics have become more accessible and widely used due to the advancement of computational power and statistical software. Many sophisticated techniques now exist for analyzing and interpreting circular data, including circular-linear regression, circular mixed-effects models, circular clustering methods, and circular data visualization tools (Cremers and Klugkist, 2018). Circular statistics have been utilized in many fields of research including neuronal activity and other biological researches such as immunology (Cremers and Klugkist, 2018; Karoly et al., 2021; Landler et al., 2019; Gregg et al., 2023). In this paper, circular statistics is used to show periodicity in physiological signals across a natural cycle by analyzing the changes in signal/feature amplitude.

Furthermore, different methods were introduced to test for a deviation from uniformity in circular data, such as the Rayleigh, Rao's spacing, V, Omnibus and Watson-Williams tests. Note that this study focuses primarily on the Rayleigh test, which evaluates periodicity and non-uniform patterns in data. Additionally, the Watson-William's test was used to compare ovulating and non-ovulating cycles. The alternative options (Rao, Omnibus, and V tests) were also examined for a comparative discussion.

In this study, the conventional LH-testing method for tracking ovulation was combined with that of previously mentioned physiological signals (HR, HRV and skin temperature) in addition to the electrodermal activity (EDA), among ovulating and non-ovulating women. Several features from temperature, HR, inter-beat interval (IBI), and EDA were extracted. The features to best represent a biphasic pattern (identifiable in ovulating cycles) were selected for the subsequent circular statistical analysis. Finally, we also report accuracies for predicting physiological signals using an Autoregressive Integrated Moving Average (ARIMA) model, which is, in essence, a regression algorithm specifically tailored for time series forecasting. Throughout the text, the cycles are labeled as either ovulating or non-ovulating based on whether the subject's LH test during that cycle showed positive or negative, respectively.

The paper is organized as follows: Materials and methods provides detail on data collection and processing techniques, as well as, analytical methods, including circular statistics, and ARIMA modeling. In the Results section, features with a noticeable non-uniform pattern are determined, and a comparison between ovulating and non-ovulating cycles has been performed. Finally, the Discussion and Conclusion sections are presented.

2 Materials and methods

To examine the relationship between changes in physiological signals, and menstrual cycles, fifteen healthy female individuals were recruited to wear a research-grade wearable wristband (two to four menstrual cycles) while tracking their ovulation. Collected data was processed, and the best features that show deviation from uniformity in ovulating subjects was determined using circular statistic test, Rayleigh. Additionally, an ARIMA model was designed to predict physiological data for the following day. The following sub-sections describe in detail the subject recruitment process, physiological data, and data analysis techniques.

2.1 Ambulatory data collection

The Empatica E4 wristband (Empatica Inc., Boston MA) was used to continuously collect physiological data across the menstrual cycles. An Institutional Review Board protocol (IRB 18-00628) has been approved to collect data from female subjects (18–40 years old) at the University of North Florida. Fifteen subjects were recruited to wear the wristband for at least one menstrual cycle, during which each subject marked their menstruation and ovulation days on a calendar. To detect ovulation, each subject was asked to utilize urine test strips, starting the day after menstruation for 2 weeks, or until the test strip depicted a positive result. Subjects who had history of pregnancy within 6 months prior, breastfeeding, working at night shifts, frequently flying across time zones or experiencing sleep disorders were excluded from the study.

TABLE 1 Demographic information and characteristics of participants included in the analysis.

	Ovulating (N = 12, 31 cycles)	Non-ovulating (N = 3, 8 cycles)
Age, mean	24.25 ± 2.93	24.33 ± 2.31
Age groups, years		
18–24	8	2
25–30	3	1
31–35	1	
BMI, mean (lb/in^2)	27.07 ± 4.47	25.744 ± 1.48
BMI groups		
<18.5		
18.5–24.9	6	1
25.0–29.9	3	2
≥30.0	3	
Duration of menstruation, mean (days)	5.48 ± 0.78	5.13 ± 1.25
Duration of cycle, mean (days)	28.87 ± 4.19	28.36 ± 1.77

In total, data was collected from 46 cycles. Removed were cycles missing more than 5 consecutive days (5 cycles) as well as cycles in which the subject had major sickness (2 cycles) that could affect temperature and heart rate, resulting in statistical analysis of 39 cycles in total; 31 of these cycles had confirmed ovulation, while 8 cycles were absent of ovulation. Non-ovulating subjects were on a birth control, however, they were asked to use the urine test to confirm this. A demographic breakdown of the participating subjects can be found in Table 1.

2.2 Physiological signals

The Empatica E4 Wristband collects several physiological signals: electrodermal activity (EDA), temperature, inter-beat-interval (IBI), accelerometer, and blood volume pulse (BVP). The Blood Volume Pulse (BVP) is the primary output from the PPG data (Torres et al., 2016), which is obtained by optical sensors (Fortino and Giampà, 2010). HR and IBI are derived from BVP, where IBI is the time interval between individual heartbeats and is computed from detecting peaks of the BVP and its variability defines the HRV (Shaffer and Ginsberg, 2017). Statistical features were extracted from each individual signal. From IBI, autocorrelation, LF and HF powers in frequency bands of 0.04–0.15 Hz, and 0.15–0.4 Hz respectively, and normalized LF power were calculated (Tenan et al., 2014).

Electrodermal Activity (EDA), usually measured in μS (micro Siemens), refers to the electrical changes measured on the epidermis, which arises when it receives innervating signals, such as stress, from the brain (Benedek and Kaernbach, 2010; Djawad et al., 2019; Shukla et al., 2019; Zangróniz et al., 2017). By applying a low voltage, EDA sensors can non-invasively measure skin conductance (Djawad

et al., 2019). The skin conductance is characterized into two types: Tonic Skin Conductance Level (SCL), and rapid Phasic Skin Conductance Response (SCR) (Braithwaite et al., 2013). The Tonic SCL is generally considered to be the slowly varying component of skin conductance. The other component, SCR, is the faster changing element (Braithwaite et al., 2013; Benedek and Kaernbach, 2010) and represented by bursts of peaks. Numerous studies (e.g. (Zangróniz et al., 2017; Benedek and Kaernbach, 2010; Egan et al., 2016)) have shown that there is a direct correlation between physiological signals, such as heart rate, and EDA.

From EDA, statistical features, area under both phasic and tonic components, signal magnitude area (SMA) in both frequency and time domains, SCR peak count, SCR peak width as well as normalized power in frequency ranges between 0.1 and 0.5 in steps of 0.1, were calculated. The SMA in frequency domain, was calculated by taking the sum of the absolute value of the Fourier Transform of the phasic component.

2.3 Minimizing artifacts

The challenging aspect with wearable devices is the quality of the data recorded, which becomes more important in ambulatory settings and long-term data collection. Several algorithms were introduced to remove data epochs corrupted by artifacts using signal quality metrics (Böttcher et al., 2022; Nasseri et al., 2020). It was shown that the quality of the signals were higher at night than during the day (Böttcher et al., 2022), because motion artifacts are the main source of data corruption.

Therefore, some studies focus only on analyzing data recorded during the night, or sleep times. In Shilaih et al. (2018) subjects were asked to wear an Ava bracelet during sleep to continuously record wrist skin temperature (WST). The Ava bracelet provided one measurement every 10-s. To avoid variation in temperature induced by the sleep onset and waking up, the first 90 and the last 30-min of each night's data were excluded. To remove artificial fluctuations, the temperature signal was smoothed before statistical analysis. In another study, nocturnal data was collected using the Oura ring from 10-pm to 8-am. To remove the data fluctuations, a moving average filter with a length of 17-min was applied to data (Maijala et al., 2019).

Although data in this study was collected from subjects wearing the device the whole time, to minimize the effects of artifacts, sleep data was extracted during no-movement (or minimum movement) hours by implementing a simple algorithm to detect changes in hand angle using accelerometry data (van Hees et al., 2018). A threshold was determined to detect the no-movement hours, within which actual sleep hours were identified via HR thresholding. The next section describes the sleep data extraction in more detail.

2.4 Data preprocessing

The heuristic algorithm applied to each subject's cycle is illustrated in Figure 2 after removing cycles missing more than 5 days or cycles in which the subject had a major sickness. There is variability among what a regular menstrual cycle length range should be; however, many studies fall in the range from

FIGURE 2
Steps of the heuristic algorithm for processing and plotting each physiological signal.
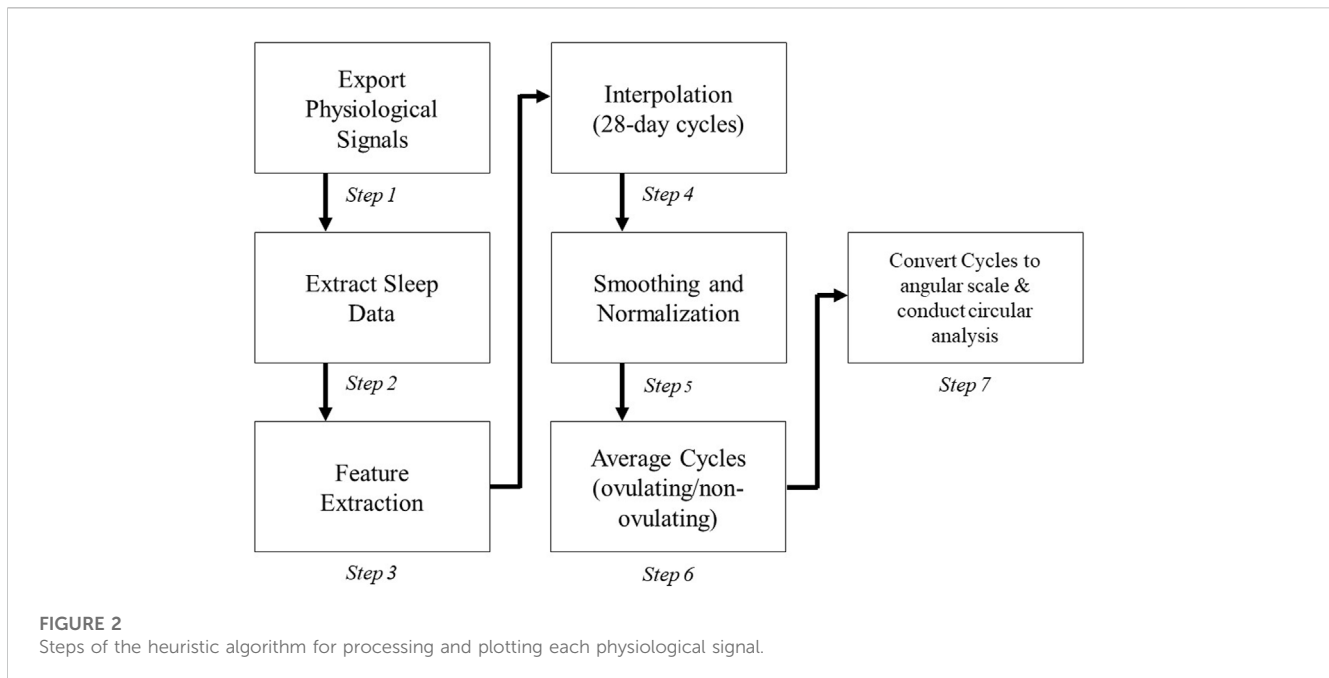
25–35 days for healthy pre-menopausal women (Campbell et al., 2021; Schmalenberger et al., 2021; Fehring et al., 2006). The cycles outside of this range with a positive luteinizing hormone test (indicating ovulation) were not removed from the study. Collected cycles in our study are in the range of 22–38 days with mean cycle duration of 28.87 ± 4.19 days (Table 1).

Although data was collected continuously, only the data during the subjects' sleep was extracted to calculate daily values for each feature (Steps 1-3 in Figure 2) due to the inactivity of the subject and less motion artifacts (van Hees et al., 2018). Sleep hours were determined based on changes in the angle of the hand derived from accelerometry data: first calculated was the 5-min rolling median of the absolute differences between consecutive 5-s mean of the z-angle of data recorded during 8-pm to next day 10-am; then calculated was the 10th percentile of this data multiplied by 15 to define a threshold (van Hees et al., 2018); and finally the time of data less than this threshold specified sleep hours. Other criteria were also considered, including heart rate thresholds (< 90 bpm) and sleep duration (> 1-h). Lastly, daily average values were calculated from features extracted from data recorded within sleep hours.
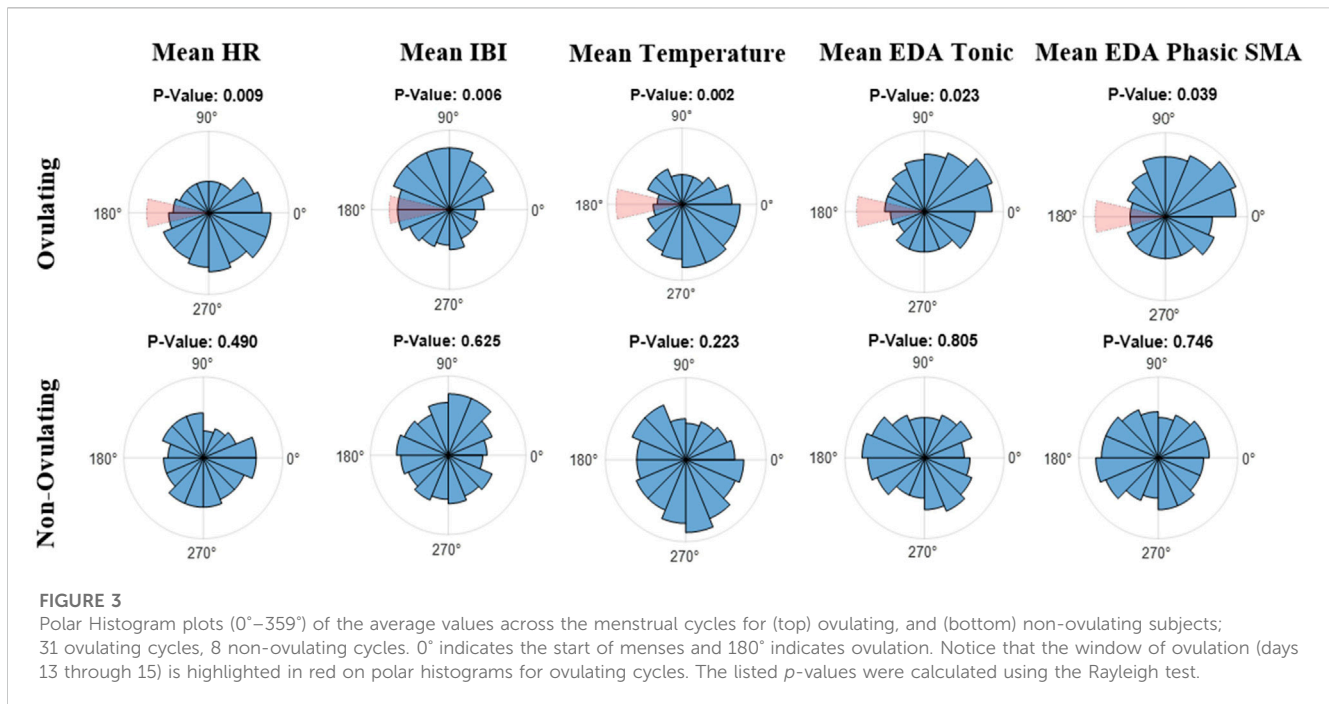
In the next 2 steps (Step 4–5) cycles were interpolated (i.e., up/down sampled) to occupy a 28-day period to account for inconsistency in the cycle ranges. Data was normalized to the range of 1–10 after smoothing it by applying a causal moving average filter with length of 4. Finally, subjects' cycles were averaged and converted from a linear scale to an angular scale (Steps 6–7).

2.5 Statistical analysis

To conduct statistical analysis, the phases of each cycle should be determined first. The menstrual cycle occurs in three main phases: follicular, ovulation, and luteal. Each interpolated 28-day cycle can be further classified into the five following phases: 1) menses (days

1–5), 2) mid-follicular phase (days 7–11), 3) ovulation phase (days 13–15), 4) mid-luteal phase (days 19–23) and 5) late-luteal phase (days 24–28). Menses was determined using the average menstruation duration listed in Table 1. The follicular phase is defined as the start of menstruation up until the day of ovulation (Fehring et al., 2006; Schmalenberger et al., 2021). Thus, the mid-follicular phase was determined to be 2 days post menstruation up to 2 days prior to ovulation. For an average 28-day cycle, the transition from the follicular phase to the luteal phase is known as ovulation and generally occurs on day 14 of the cycle (Prior et al., 2015). Therefore, the ovulation phase was defined to be days 13–15 to ensure inclusivity of the day of ovulation. The mid-luteal phase is characterized by constant low LH and FSH levels with elevated levels of progesterone (Figure 1B). Therefore, the mid-luteal phase was defined as days 19–23. Finally, the late-luteal phase starts when progesterone decreases and ends the day before menses (days 24–28). To show the significant difference between the different combinations of phases, a paired *t*-test was used.

Lastly, data was mapped on an angular scale (0°–359°) using Matlab's CircStat toolbox (Berens, 2009). In reference to the plots in Figure 3, 0° indicates the start of menses, or the first day of the subjects' cycles, 180° marks ovulation, and 359° marks the end of the LH phase, and consequently the end of the 28-day cycle. Analyzed were both subject data on a per cycle basis as well as the combined average of all cycles. Figure 3 refers to the overall average of ovulating and non-ovulating cycles, respectively. The Rayleigh test was utilized for both ovulating and non-ovulating groups to detect the existence of a unimodal deviation from uniformity. A small p ($p < 0.05$) indicates a significant egress from uniformity and thus rejects the null hypothesis of uniform distribution (Berens, 2009). Rejections of the null hypothesis is an indication of a non-uniform pattern and thus represents ovulation. The Watson-Williams test, or the circular analogue of the two-sample *t*-test (Berens, 2009), was used to compare the two groups, namely, ovulating and non-ovulating cycles. This test assumes underlying

FIGURE 3
Polar Histogram plots (0°−359°) of the average values across the menstrual cycles for (top) ovulating, and (bottom) non-ovulating subjects; 31 ovulating cycles, 8 non-ovulating cycles. 0° indicates the start of menses and 180° indicates ovulation. Notice that the window of ovulation (days 13 through 15) is highlighted in red on polar histograms for ovulating cycles. The listed *p*-values were calculated using the Rayleigh test.

von Mises distributions with equal concentration parameter and evaluates whether the mean directions of two or more groups are identical (Berens, 2009). Finally, to show the significant difference between the different combinations of phases, including menses, mid-follicular, ovulation, mid-luteal and late-luteal, a paired *t*-test was used.

2.6 Autoregressive integrated moving average (ARIMA) model

To predict physiological data for the following day, an autoregressive integrated moving average (ARIMA) algorithm was employed, which is well-suited for time series analysis. ARIMA models consist of three components: autoregressive (AR), integrated (I), and moving average (MA) terms. The AR term captures the relationship between a variable and its past values. The MA term captures the relationship between a variable and its past error terms, and the I part introduces differencing to make the time series stationary.

To determine the appropriate ARIMA model, first a moving average filter was applied to smooth out fluctuations, and then Augmented Dickey Fuller (ADF) test was used to confirm that data is stationarity ($p < 0.05$), thus the differencing parameter (known as *d*) was set to zero. To define the rest of the parameters the Autocorrelation Function (ACF), and Partial Autocorrelation Function (PACF) were used. Where the ACF measures the correlation between a time series and its lags at different intervals, and the PACF measures the correlation between a time series and its lags while controlling for shorter lags. Observing ACF and PACF plots (using training data for each subject individually), and after identifying significant spikes at certain lags (Kotu and Deshpande, 2018), the AR and MA terms for the ARIMA model was determined to be 2 and 3 (known as *p* and *q*). These values were

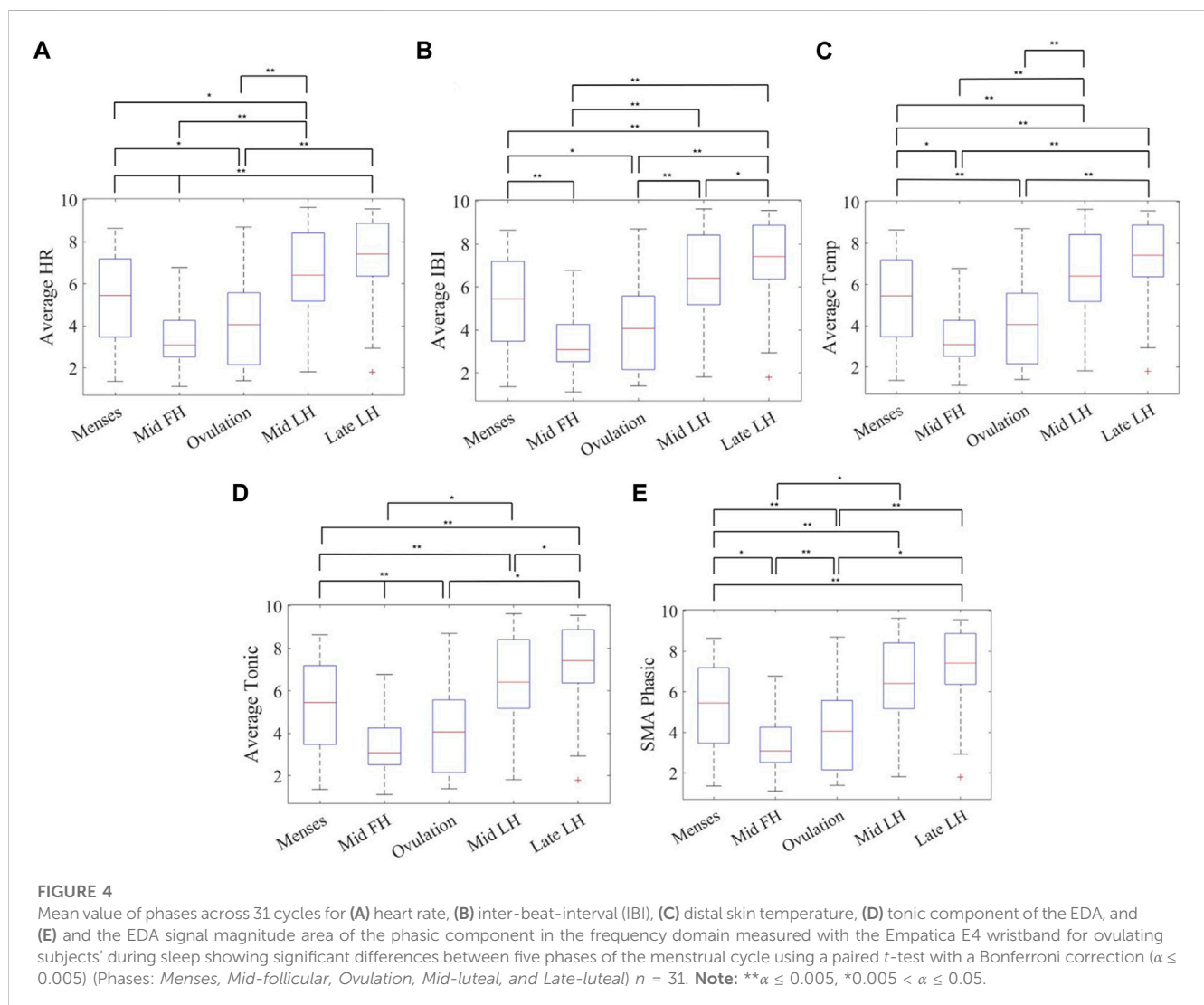selected since they were identifying significant spikes for majority of the subjects.

Here only the cycles with missing data (5 cycles) were removed which resulted in a total of 41 ovulating and non-ovulating cycles: more specifically, data included one subject with one cycle, 3 subjects with only 2 cycles each, 10 subjects with 3 cycles each and one subject with 4 cycles. Data from the subject with only one cycle was removed as well, resulting in 40 cycles. From mean values of temperature, HR, IBI and tonic EDA, the last cycle was kept for test while the rest (2 first cycles for 10 subjects, one first cycle for 3 subjects, and 3 first cycles for one subject) were used for initial training. It should be noted that each subject was fitted separately. The algorithm was re-trained every day by adding the new daily measure to training dataset, to predict the next day value. By analyzing patterns and trends in the data, the algorithm can identify potential changes and predict what is likely to happen in the next day.

3 Results

Using the Rayleigh test, in ovulating subjects, a non-uniform pattern was observed with a significant periodicity ($p < 0.05$) in mean temperature, HR, IBI, EDA tonic component, and SMA of the EDA phasic component in frequency domain among the extracted features reported in Section 2.2. Similar to the BBT chart in Figure 1A, there is a noticeable dip in amplitude of the polar histogram around 180° in the ovulating mean temperature chart in Figure 3 (top), indicating ovulation and representing a non-uniform distribution with a biphasic pattern ($p < 0.05$). The biphasic pattern was observed in mean EDA tonic and SMA of EDA phasic as well, although not as distinct as the mean temperature pattern. In contrast, non-ovulating cycles displayed a more uniform distribution in all selected features ($p > 0.05$).

TABLE 2 Comparison of ovulating and non-ovulating cycles using the Watson-Williams test (Berens, 2009). This displays the differences between ovulating and non-ovulating averages from the following features: average HR, IBI, temperature, EDA skin conductance level, and SMA of the EDA phasic component in the frequency domain.

Feature	Test-statistic (F)	*p*-value
Mean Heart Rate	0.08495	0.770911343
Mean of IBI	5.10189	0.024588662
Mean Temperature	4.42631	0.036160184
EDA Signal Magnitude Area of Phasic Component	13.5693	0.000284669
Mean EDA Skin Conductance Level (Tonic Component)	20.6076	8.79E-06

FIGURE 4

Mean value of phases across 31 cycles for **(A)** heart rate, **(B)** inter-beat-interval (IBI), **(C)** distal skin temperature, **(D)** tonic component of the EDA, and **(E)** and the EDA signal magnitude area of the phasic component in the frequency domain measured with the Empatica E4 wristband for ovulating subjects' during sleep showing significant differences between five phases of the menstrual cycle using a paired *t*-test with a Bonferroni correction ($\alpha \leq$ 0.005) (Phases: *Menses, Mid-follicular, Ovulation, Mid-luteal, and Late-luteal*) n = 31. **Note:** **$\alpha \leq$ 0.005, *0.005 < $\alpha \leq$ 0.05.

Using the Watson-Williams test, a comparison was made between ovulating and non-ovulating groups. Unlike mean temperature, IBI and EDA, there was no significant difference (p > 0.05) between ovulating and non-ovulating cycle groups in mean HR (Table 2). The results of the tonic component were particularly interesting. Notice that in ovulating cycles (Figure 3), ovulation had a lower amplitude. In contrast, the associated window

of ovulation for non-ovulating cycles had an elevated amplitude, albeit not significant compared to the rest of phases.

Comparison between combinations of phases including menses, mid-follicular, ovulation, mid-luteal and late-luteal, was performed using paired *t*-test after extracting the mean values of each phase from the smoothed and normalized data. In Figure 4, and outlined in Supplementary Table S1, ovulating cycles show significant

differences between late-luteal phase and follicular phases (menses, mid-follicular) for mean temperature, HR and IBI as well as between late-luteal and menses for EDA mean tonic and SMA using a paired t-test with Bonferroni correction (α of 0.005). There is a significant difference between ovulation and mid-luteal phases in mean HR, IBI and temperature. No significant difference is observed between the mid-follicular phase and ovulation, except for EDA mean tonic and SMA.

Further, the ARIMA model was used to accurately predict the mean temperature, HR, IBI and EDA tonic of the following day (ovulating and non-ovulating), which resulted in the average root mean square error (RMSE) of 0.13 ± 0.07 (C°), 1.31 ± 0.34 (bpm), 0.016 ± 0.005 (s), and 0.17 ± 0.17 (μS) respectively. Average RMSE across ovulating cycles for mean temperature, HR, IBI, and EDA tonic are 0.12 ± 0.03 (C°), 1.36 ± 0.35 (bpm), 0.017 ± 0.005 (s) and 0.15 ± 0.17 0.06 (μS), respectively. The actual and predicated average daily values are shown in Supplementary Figure S1 for the third cycle of an ovulating subject with length of 26 days. The predicted curve is shown to follow the same pattern of the actual curve. These relatively low RMSE values show that ARIMA can be used to predict physiological signals based on previous data. This suggests wearable devices are a promising technology for predicting menstrual cycle phases using an appropriate ARIMA model.

4 Discussion

4.1 Principal findings

The purpose of the study was to implement circular statistics to investigate changes in physiological features during the natural menstrual cycle in women of reproductive age. Due to the periodic nature of the data used in this study, signals were converted to an angular scale and presented in degrees on a circular plot, which is different from commonly used linear techniques. On a circle 0° and 360° demonstrate the same direction, however, on a linear scale those points represent opposite ends of a scale, and much further apart (Cremers and Klugkist, 2018). Circular statistics gave us a tool to visualize the data in a different way.

Data from a wearable device (Empatica E4 wristband), combined with self-reported ovulation detection (LH-test strip) and a calendar with the mark of the start and end of menses were used to determine features representing significant phase shifts for ovulating cycles. Ovulating and non-ovulating cycles were also compared using an angular approach. Further, to determine the significance in deviation from uniformity in each feature, the Rayleigh test was used. The Rayleigh test was selected by observing the pattern in data; the distribution with one clear peak was observed in features after the occurrence of the nadir (lowest point) as in Figure 1A. However, the Rayleigh test generally assumes that the data is distributed normally (Karoly et al., 2018). To check uniformity, other statistical tests might also be considered such as the Rao's spacing test, V test and Omnibus. Rao's test was not a suitable option for this study because it is more applicable to data, that is, neither unimodal nor axially bimodal (Berens, 2009) and it did not confirm any significant phase shift in ovulating or non-ovulating subjects. The V test showed similar results as the Rayleigh

test except for mean IBI. The V test for circular uniformity is similar to the Rayleigh test except that the mean direction of data has to be known before analyzing the data (Berens, 2009). The Omnibus test (Hodges-Ajne test) confirmed circular uniformity in ovulating subjects in mean temperature only. This test detects general deviations from uniformity but with less statistical power and without assumptions about the underlying distribution (Berens, 2009). In other words, the Omnibus test is used to check whether there is any circular pattern or non-random structure in the data. Although different tests were introduced to evaluate distribution uniformity, for unimodal distributions, the Rayleigh and V tests are more powerful than the Omnibus test, and among these two, the Rayleigh test is best recommended for unimodal departures from uniformity. However, in the multimodal case, the alternative tests should be considered (Landler et al., 2019). The Rayleigh test is suitable for unimodal distribution, but it might not be the best test to confirm biphasic pattern. Additionally, while both the Omnibus test and the Rayleigh test are used to assess the uniformity of circular data, the Omnibus test is more general and examines overall circular pattern, but the Rayleigh test is more focused on detecting unimodal distributions and the presence of specific orientations, or clustering, in circular data. It should be also noted that, although a biphasic pattern was observed in a few features such as mean temperature, it does not present a second significant peak to be considered as a bimodal distribution.

The findings demonstrated that wearable technology, with the ability to monitor multiple physiological features concurrently, including HR, IBI, temperature, and the EDA, can capture differences between ovulating and non-ovulating cycles. Of the many features extracted from the varying signals, only those with great significance in showing a non-uniform pattern regarding overall averages of cycles were reported. Many features had significance on a per subject, per cycle basis. One such feature, for example, was the power in 0.3–0.4 Hz band of the EDA phasic component. This feature, in particular, displayed significant phase shift in 10 out of the 12 ovulating subject averages. However, cycle averages indicated a more uniform distribution (p > 0.05). It was also interesting that a significant change was not observed across the menstrual cycle in LF and HF components of IBI despite findings in Vishrutha et al. (2012).

Additionally, the selected features were used in a regression model, specifically an Autoregressive Integrated Moving Average, to predict the physiological signals. It was shown that the ARIMA model accurately predicted the physiological values for the testing set, as evidenced by the relatively low value of the RMSE. Not a significant difference has been observed between ovulating and non-ovulating cycles in terms of RMSE. It is worth mentioning that data from non-ovulating cycles could be periodic but not biphasic. However, more non-ovulating cycles are needed to draw a clear conclusion.

The result from this study has many implications in reference to monitoring and predicting menstrual phases. For example, a reproductive-aged woman taking medication while trying to conceive would have the ability (along with medical advice from a licensed physician) to monitor medication dosages depending on the menstrual phase. Also, people with health issues that are linked to specific menstrual phase, such as catamenial epilepsy, can benefit from knowing if their cycle is ovulatory and accordingly adjust the anti-seizure medication dosage.

4.2 Limitations

There are several study limitations that may affect the accuracy of the wearable device. This includes, but is not limited to, device position, demographic factors and sleep assessment. Thus, the possibility that the Empatica E4 wristband may exhibit different levels of accuracy across different subjects, cycles and menstrual phases must be considered. A further limitation was that blood hormonal levels were not taken. This would have provided a more accurate method for detecting not only phases but ovulation itself. Although LH-tests (urine) were used, which are 99% accurate, there is potential for false negatives if the test was used incorrectly. A further limitation concerns the relatively small sample population size of non-ovulating cycles, especially compared to ovulating case (8 non-ovulating to 31 ovulating cycles). While we concluded the non-ovulating cycles exhibit a more uniform distribution, additional data has the potential to reveal that non-ovulating cycles also present a non-uniform distribution, but no less distinct from the non-uniform distribution of ovulating cycles. A larger sample size would, not only give better results, but accurately represent the two groups better.

5 Conclusion

In conclusion, angular methods can accurately represent the biphasic and non-uniform pattern in the data collected using a wearable device for naturally occurring menstrual cycles of reproductive-aged women. There is significant difference in ovulating and non-ovulating cycles. These results can fuel algorithms, such as ARIMA, to accurately predict human physiological signals and potentially classify menstrual phases via machine learning algorithms. Thus, wearable technology is a promising tool for tracking physiological changes across the menstrual cycle.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by the University of North Florida Institutional Review Board (IRB 18-00628). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KS, GK, MT, and MN contributed in data analysis. KS, MN, PK, and BHB designed the study. KS and MN drafted the manuscript. MN and MT collected, organized and managed the dataset. GK wrote a section of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author BHB declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnetp.2023.1227228/full#supplementary-material

References

Alzueta, E., de Zambotti, M., Javitz, H., Dulai, T., Albinni, B., Simon, K. C., et al. (2022). Tracking sleep, temperature, heart rate, and daily symptoms across the menstrual cycle with the oura ring in healthy women. *International Journal of Women's Health* 14, 491–503. doi:10.2147/IJWH.S341917

Benedek, M., and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods* 190, 80–91. doi:10.1016/j.jneumeth.2010.04.028

Berens, P. (2009). Circstat: a matlab toolbox for circular statistics. *Journal of statistical software* 31, 1–21. doi:10.18637/jss.v031.i10

Böttcher, S., Vieluf, S., Bruno, E., Joseph, B., Epitashvili, N., Biondi, A., et al. (2022). Data quality evaluation in wearable monitoring. *Scientific reports* 12, 21412. doi:10.1038/s41598-022-25949-x

Braithwaite, J. J., Watson, D. G., Jones, R., and Rowe, M. (2013). Publication recommendations for electrodermal measurements: publication standards for EDA. *Psychophysiology* 49, 1017–1034. doi:10.1111/j.1469-8986.2012.01384.x

Bull, J. R., Rowland, S. P., Scherwitzl, E. B., Scherwitzl, R., Danielsson, K. G., and Harper, J. (2019). Real-world menstrual cycle characteristics of more than

600,000 menstrual cycles. *NPJ digital medicine* 2, 83. doi:10.1038/s41746-019-0152-7

Campbell, L. R., Scalise, A. L., DiBenedictis, B. T., and Mahalingaiah, S. (2021). Menstrual cycle length and modern living: a review. *Current opinion in endocrinology, diabetes, and obesity* 28, 566–573. doi:10.1097/MED.0000000000000681

Coyne, M. D., Kesick, C. M., Doherty, T. J., Kolka, M. A., and Stephenson, L. A. (2000). Circadian rhythm changes in core temperature over the menstrual cycle: method for noninvasive monitoring. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 279, R1316–R1320. doi:10.1152/ajpregu.2000.279.4.R1316

Cremers, J., and Klugkist, I. (2018). One direction? a tutorial for circular data analysis using r with examples in cognitive psychology. *Frontiers in psychology* 9, 2040. doi:10.3389/fpsyg.2018.02040

Djawad, Y. A., SaharuddinJaya, H., Ridwansyah, and Thayeb, M. (2019). "Proficiency test analysis of a simple electro-dermal activity measurement technique for measuring an emotional task," in *AIP Conference Proceedings* (AIP Publishing LLC), 2155.020050

Egan, D., Brennan, S., Barrett, J., Qiao, Y., Timmerer, C., and Murray, N. (2016). "An evaluation of heart rate and electrodermal activity as an objective qoe evaluation method for immersive virtual reality environments," in *2016 eighth international conference on quality of multimedia experience (QoMEX)* (IEEE), 1–6.

Fehring, R. J., Schneider, M., and Raviele, K. (2006). Variability in the phases of the menstrual cycle. *Journal of Obstetric, Gynecologic & Neonatal Nursing* 35, 376–384. doi:10.1111/j.1552-6909.2006.00051.x

Fisher, N. I. (1995). *Statistical analysis of circular data*. cambridge university press.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character* 222, 309–368.

Fortino, G., and Giampà, V. (2010). "Ppg-based methods for non invasive and continuous blood pressure measurement: an overview and development issues in body sensor networks," in *2010 IEEE International Workshop on Medical Measurements and Applications* (IEEE), 10–13.

Goodale, B. M., Shilaih, M., Falco, L., Dammeier, F., Hamvas, G., and Leeners, B. (2019). Wearable sensors reveal menses-driven changes in physiology and enable prediction of the fertile window: observational study. *Journal of medical Internet research* 21, e13404. doi:10.2196/13404

Gregg, N. M., Pal Attia, T., Nasseri, M., Joseph, B., Karoly, P., Cui, J., et al. (2023). Seizure occurrence is linked to multiday cycles in diverse physiological signals. *Epilepsia* 64, 1627–1639. doi:10.1111/epi.17607

Hamper, J. (2020). 'catching ovulation': exploring women's use of fertility tracking apps as a reproductive technology. *Body & Society* 26, 3–30. doi:10.1177/1357034x19898259

Karoly, P. J., Freestone, D. R., Boston, R., Grayden, D. B., Himes, D., Leyde, K., et al. (2016). Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. *Brain* 139, 1066–1078. doi:10.1093/brain/aww019

Karoly, P. J., Goldenholz, D. M., Freestone, D. R., Moss, R. E., Grayden, D. B., Theodore, W. H., et al. (2018). Circadian and circaseptan rhythms in human epilepsy: a retrospective cohort study. *The Lancet Neurology* 17, 977–985. doi:10.1016/S1474-4422(18)30274-6

Karoly, P. J., Stirling, R. E., Freestone, D. R., Nurse, E. S., Maturana, M. I., Halliday, A. J., et al. (2021). Multiday cycles of heart rate are associated with seizure likelihood: an observational cohort study. *EBioMedicine* 72, 103619. doi:10.1016/j.ebiom.2021.103619

Kiranmai, T. S., and Lakshmi, P. (2021). A comprehensive and systematic literature review of computational intelligence algorithms to diagnose and predict female infertility. *Annals of the Romanian Society for Cell Biology*, 5926–5943.

Kotu, V., and Deshpande, B. (2018). *Data science: concepts and practice (Morgan Kaufmann)*.

Landler, L., Ruxton, G. D., and Malkemper, E. P. (2019). Circular statistics meets practical limitations: a simulation-based rao's spacing test for non-continuous data. *Movement Ecology* 7, 15–5. doi:10.1186/s40462-019-0160-x

Maijala, A., Kinnunen, H., Koskimäki, H., Jämsä, T., and Kangas, M. (2019). Nocturnal finger skin temperature in menstrual cycle tracking: ambulatory pilot study using a wearable oura ring. *BMC Women's Health* 19, 150–10. doi:10.1186/s12905-019-0844-9

Martinez, A. R., van Hooff, M. H., Schoute, E., van der Meer, M., Broekmans, F. J., and Hompes, P. G. (1992). The reliability, acceptability and applications of basal body temperature (bbt) records in the diagnosis and treatment of infertility. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 47, 121–127. doi:10.1016/0028-2243(92)90041-v

Mathworks (2022). *ttest, matlab version: 9.13.0 (r2022b)*.

Messinis, I. E., Messini, C. I., and Dafopoulos, K. (2010). The role of gonadotropins in the follicular phase. *Annals of the New York Academy of Sciences* 1205, 5–11. doi:10.1111/j.1749-6632.2010.05660.x

Nasseri, M., Nurse, E., Glasstetter, M., Böttcher, S., Gregg, N. M., Laks Nandakumar, A., et al. (2020). Signal quality and patient experience with wearable devices for epilepsy management. *Epilepsia* 61, S25–S35. doi:10.1111/epi.16527

Pewsey, A., and García-Portugués, E. (2021). Recent advances in directional statistics. *Test* 30, 1–58. doi:10.1007/s11749-021-00759-x

Prior, J. C., Naess, M., Langhammer, A., and Forsmo, S. (2015). Ovulation prevalence in women with spontaneous normal-length menstrual cycles–a population-based cohort from hunt3, norway. *PloS one* 10, e0134473. doi:10.1371/journal.pone.0134473

San Roman, G., Long, C. A., Reshef, E., Dodds, W., and Gast, M. J. (1995). Monitoring the ovulation induction cycle. *American journal of obstetrics and gynecology* 172, 785–788. doi:10.1016/0002-9378(95)90155-8

Schmalenberger, K. M., Eisenlohr-Moul, T. A., Jarczok, M. N., Eckstein, M., Schneider, E., Brenner, I. G., et al. (2020). Menstrual cycle changes in vagally-mediated heart rate variability are associated with progesterone: evidence from two within-person studies. *Journal of clinical medicine* 9, 617. doi:10.3390/jcm9030617

Schmalenberger, K. M., Tauseef, H. A., Barone, J. C., Owens, S. A., Lieberman, L., Jarczok, M. N., et al. (2021). How to study the menstrual cycle: practical tools and recommendations. *Psychoneuroendocrinology* 123, 104895. doi:10.1016/j.psyneuen.2020.104895

Setton, R., Tierney, C., and Tsai, T. (2016). The accuracy of web sites and cellular phone applications in predicting the fertile window. *Obstetrics & Gynecology* 128, 58–63. doi:10.1097/AOG.0000000000001341

Shaffer, F., and Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health* 5, 258. doi:10.3389/fpubh.2017.00258

Shilaih, M., Goodale, B. M., Falco, L., Kübler, F., De Clerck, V., and Leeners, B. (2018). Modern fertility awareness methods: wrist wearables capture the changes in temperature associated with the menstrual cycle. *Bioscience reports* 38, BSR20171279. doi:10.1042/BSR20171279

Shukla, J., Barreda-Angeles, M., Oliver, J., Nandi, G. C., and Puig, D. (2019). Feature extraction and selection for emotion recognition from electrodermal activity. *IEEE Transactions on Affective Computing* 12, 857–869. doi:10.1109/taffc.2019.2901673

Sinex, J. E. (1999). Pulse oximetry: principles and limitations. *The American journal of emergency medicine* 17, 59–67. doi:10.1016/s0735-6757(99)90019-0

Su, H.-W., Yi, Y.-C., Wei, T.-Y., Chang, T.-C., and Cheng, C.-M. (2017). Detection of ovulation, a review of currently available methods. *Bioengineering & translational medicine* 2, 238–246. doi:10.1002/btm2.10058

Tenan, M. S., Brothers, R. M., Tweedell, A. J., Hackney, A. C., and Griffin, L. (2014). Changes in resting heart rate variability across the menstrual cycle. *Psychophysiology* 51, 996–1004. doi:10.1111/psyp.12250

Torres, J. M. M., Ghosh, A., Stepanov, E. A., and Riccardi, G. (2016). "Heal-t: an efficient ppg-based heart-rate and ibi estimation method during physical exercise," in *2016 24th European signal processing conference (EUSIPCO)* (IEEE), 1438–1442.

van Hees, V. T., Sabia, S., Jones, S. E., Wood, A. R., Anderson, K. N., Kivimäki, M., et al. (2018). Estimating sleep parameters using an accelerometer without sleep diary. *Scientific reports* 8, 12975. doi:10.1038/s41598-018-31266-z

Vishrutha, K., Harini, N., Ganaraja, B., Pavanchand, A., and Veliath, S. (2012). *A study of cardiac autonomic control and pulmonary functions in different phases of menstrual cycle*.

Yu, J.-L., Su, Y.-F., Zhang, C., Jin, L., Lin, X.-H., Chen, L.-T., et al. (2022). Tracking of menstrual cycles and prediction of the fertile window via measurements of basal body temperature and heart rate as well as machine-learning algorithms. *Reproductive Biology and Endocrinology* 20, 118–12. doi:10.1186/s12958-022-00993-4

Zangróniz, R., Martínez-Rodrigo, A., Pastor, J. M., López, M. T., and Fernández-Caballero, A. (2017). Electrodermal activity sensor for classification of calm/distress condition. *Sensors* 17, 2324. doi:10.3390/s17102324