2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)

# A Rhythm Analysis-Based Model to Predict Sedentary Behaviors

Qian He
*Department of Computer Science*
*Worcester Polytechnic Institute*
*Worcester, Massachusetts 01609–2280*
*Email: qhe@cs.wpi.edu*

Emmanuel O. Agu
*Department of Computer Science*
*Worcester Polytechnic Institute*
*Worcester, Massachusetts 01609–2280*
*Email: emmanuel@cs.wpi.edu*

***Abstract*—Sedentary behaviors such as sitting and watching TV are ubiquitous in modern societies. Increases in sedentary time have been linked with an increased risk of obesity, diabetes, cardiovascular disease, and all-cause mortality. While smartphones and wearables can now detect sedentary user behaviors, few computational models exist for predicting when they will occur in future. In this paper, we propose a lightweight model to predict future sedentary behaviors, facilitating prevention rather than reactive interventions. Our models are based on the concept of rhythm analysis, an idea proposed by Lefebvre, which postulates that many human behaviors, the use of public spaces, and many phenomena all follow natural rhythms. Our work focuses on detecting the prevailing rhythms of sedentary behaviors and modeling the cyclical rhythm and linear rhythm in Lefebvre's philosophy using periodic functions (history-free) and linear functions (history-dependent) respectively. A person who lies on his couch at the same time every day is an example of a cyclical rhythm, while a person who lies down in exhaustion after vigorous exercise is an example of a linear rhythm. Our preliminary results from analyzing an existing dataset clearly show that rhythmical sedentary patterns do exist. Cyclical rhythms are more common than linear rhythms, and half-day rhythms, daily rhythms, weekly rhythms, and bi-weekly rhythms are clearly observed in a test dataset.**

## 1. Introduction

Sedentary behaviors, defined as waking behaviors with an energy expenditure $\leq$ 1.5 METs (Metabolic Equivalent of Tasks) while in a sitting or reclining posture [1], are ubiquitous in modern societies. Examples such as sitting while using computers, watching TV, playing video games, and commuting are common in daily modern lives. Sedentary behaviors have been associated with a 112% increase in the risk of diabetes, 147% increase in cardiovascular disease, 90% increase in cardiovascular mortality, and 49% increase in all-cause mortality [2].

Sedentary behaviors caused by screen-based occupations (e.g., desk jobs) and entertainment (e.g., watching TV) in adults [3], [4], college students [5], [6], and children [7], [8] have previously been studied. Many of these prior studies have focused on post-analysis of past sedentary behaviors to determine their causes. Predicting future sedentary behaviors, however, remains an open unsolved problem. The ability to anticipate and predict future sedentary behaviors will facilitate a transformation from health behavior intervention to health behavior prevention, which will likely be more effective [9].

We propose predictive models for sedentary behaviors based on classic rhythm analysis [10], which contends that rhythms are found in many areas of human life including work schedules, seasonal patterns of influenza, the body's hunger, shopping center crowds, night and day. The dynamic interplay of rhythms yields important components of the everyday contexts of people's lives, increasing or decreasing the effort required to perform everyday activities. We hypothesize that such rhythms generally exist in many *Activities of Daily Living (ADL)* [11] and specifically in sedentary behaviors. This hypothesis is supported by the work of Rantala and Valtonen, who found rhythmic patterns in the sleep patterns of tourists [12]. McQuoid [13] also proposes that self-management of chronic ailments might be easier if patients form habits that do not conflict with rhythms already existing in their lives. In this paper, we focus on detecting rhythms of sedentary behaviors in order to predict future occurrence.

While prior medical research has established the existence of rhythms in health-related behaviors such as the circadian rhythms [14], which influence human sleep cycles and may cause mood and mental health disorders when disrupted [15], very few computational models have been proposed to capture such rhythms. We propose two canonical types of models to capture sedentary rhythms: (1) a history-free predictive model, which uses a frequency domain algorithm to predict future sedentary behaviors with cyclical rhythms, and (2) a history-dependent model, which uses the AutoRegressive (AR) model to predict future sedentary behaviors as a function of recent past behaviors (observed linear rhythms). We fit these two models to smartphone-sensed user activity logs in order to extract rhythmic patterns of sedentary behaviors and predict future sedentary behaviors at various timescales. Our results have broad impact since such activity logs can now be generated by many physical activity trackers (e.g., Fitbit) and smartphone apps (e.g., Google Fit [16]).

978-1-5090-4722-2/17 $31.00 © 2017 IEEE
DOI 10.1109/CHASE.2017.122


Figure 1. The sedentary behavior prediction problem

A unified model is then proposed to combine the history-free and history-dependent models. Two algorithms are developed to build and optimize this unified model. Our preliminary results show that cyclical (periodic) rhythms are more common than linear (based on recent past behaviors) rhythms, and half-day rhythms, daily rhythms, weekly rhythms, and bi-weekly rhythms are clearly observed in a test dataset [17]. Our application of rhythm analysis yields a lightweight model for predicting future sedentary behaviors without requiring complex contextual information (e.g., location, temperature, and humidity) sensed with sensors and computations utilizing context, which is more suitable for wearable devices such as Fitbit.

An overview of the problem addressed by this paper as well as our overarching vision are depicted in Figure 1. The rest of this paper is organized as follows: A brief review of related work is provided in Section 2. Section 3 describes our two canonical types of predictive models as well as our unified model that combines them. In Section 4, we present the results of experiments to validate our proposed predictive models and then discuss our results in Section 5. Finally, in Section 6, we conclude by summarizing our contributions, limitations, and future work.

## 2. Background and Related Work

### 2.1. Activity Prediction

*Activity Prediction* has been researched in the context of Activities of Daily Living (ADL) in smart homes [18] and for predicting user actions within videos in computer vision [19]. Pentland and Liu [20], proposed that many human behaviors can be accurately described as a set of dynamic models (e.g., Kalman filters and Markov chains). By using their behavior modeling methodology, they demonstrated that driving-related actions can be accurately categorized soon after the actions begin in order to predict future actions.

Similarly, graph-based probabilistic activity prediction models such as Hidden Markov Models [21], Conditional Random Fields [22], and rule-based activity prediction have been suggested [23]. Applications combining ideas in activity prediction with computer vision techniques have emerged in recent years [19]. For instance, Galata *et al.* [24] combined motion capture data with a Markov model in order to predict body movements in exercise routines.

Although these proposed human behavior prediction models have shown promising results in certain scenarios, they are usually complex due to the large amount of multivariate information required for making predictions (e.g., context information such as location, environment, people's mental state and physical state). In contrast, we propose models that are context-free, utilizing only the history of user's activity, which makes our models suitable for running on resource-constrained mobile/wearable devices such as activity trackers (e.g., Fitbit).

### 2.2. Circadian and Health-Related Rhythms

Prior medical research has established the existence of rhythms in health-related behaviors. The circadian rhythm is a 24-hour cycle of physical, mental, and behavioral changes in humans, which influence humans' sleep patterns and master clocks [14]. Disruptions and abnormalities in circadian rhythms have been linked with the development of mood disorders such as bipolar disorder, major depression, and seasonal affective disorder [15]. Our work investigates whether sedentary behaviors occur in a rhythmical fashion. Specifically, we synthesize a computational model for detecting underlying rhythms of sedentary behaviors in order to predict future occurrences.

## 2.3. Rhythm Analysis

In 1930s, Lefebvre pioneered the idea that rhythms are found in many areas of human life including work schedules, seasonal patterns of influenza, the body's hunger, shopping center crowds, night and day. Lefebvre postulated that two types of rhythm exist in the daily lives of humans: (1) *Cyclical Rhythms*: repetitions with determined periods or frequencies; and (2) *Linear Rhythms*: lines, trajectories, and repetitions where future behaviors are linked with recent past behaviors [25]. A person who lies on their couch at the same time every day is an example of a cyclical rhythm, while a person who lies down in exhaustion after vigorous exercise (recent past behavior) is an example of a linear rhythm.

The rhythm analysis philosophy and methods proposed by Lefebvre have been widely used for analyzing the rhythms of urban spaces and how these rhythms influence the inhabitants living in those spaces [26]. It has also been applied in medical studies to understand the causes of difficulties experienced by people afflicted with chronic diseases participating in paid work (logistical and rhythmic space-time conflicts between managing illness and paid work) [27], to identify the sleep problems of tourists [12], and to guide the use of technological devices to encourage physically active lifestyles [28].

We apply rhythm analysis in order to predict sedentary behaviors without using contextual information. We model the two everyday life rhythms (cyclical & linear rhythms) in Lefebvre's philosophy using history-free and history-dependent predictive models respectively, which we will discuss in the next section.

# 3. Predicting Sedentary Behavior Rhythms

We start by defining ***sedentary behavior*** quantitatively as the percentage of time a person is sedentary within a given time interval. For instance, if a person sits a total of 15 minutes during the 10:00AM–10:20AM (20-minute) time period, by our definition, this person had a sedentary behavior with sedentary level of 75% ($= \frac{15min}{20min} \times 100\%$). A more detailed discussion of sedentary behaviors in practice will be discussed in the Section 4 (*Experiment*).

We propose two types of models for predicting sedentary behavior: (1) history-free predictive model, which predicts the future sedentary behavior using learned cyclical rhythms; and (2) history-dependent predictive model, which predicts the future sedentary behavior using observed linear rhythms (recent past behaviors). A unified model is then introduced to combine the history-free and history-dependent models.

## 3.1. History-Free Sedentary Behavior Prediction (HF-SBP)

Cyclical rhythms or periodic patterns are common in nature. For example, most people sleep every night. We believe that some sedentary behaviors may follow periodic patterns. If such patterns exist, the prediction task switches to that of finding the periods of all active cycles of sedentary behaviors. The time series of sedentary behavior $(y_1, y_2, \dots, y_t)$ with an underlying cyclical pattern can be modeled as:

$$Y = A\,cos(\omega t + \phi) + B \quad (1)$$

where $\omega$ is the frequency, $\phi$ is the phase shift, $A$ is the amplitude, $B$ is the vertical shift, $t$ is the time, and $Y$ is the sedentary level. Given a time index $t$ indicating a sedentary behavior in the future, we can predict this sedentary behavior $y_t$ by applying Equation 1 with $t$.

At any instant, multiple cycles may be occurring, requiring the combination (summation) of several cyclical rhythms with different frequencies, phase shifts, amplitudes, and vertical shifts (e.g., Figure 2 as an example). To generalize our history-free model, we replace Equation 1 with:

$$Y = B + \sum_{i=1}^{n} A_i cos(\omega_i t + \phi_i) \quad (2)$$

where $n$ is the number of periodic functions, and $B$ is the overall vertical shift. Equation 2 uses the superposition principle of Fourier series to combine sedentary behaviors on multiple different timescales. To estimate the parameters $A_i$, $\omega_i$, and $\phi_i$ of this model, we apply discrete Fourier transform and Cooley-Tukey algorithm [29].

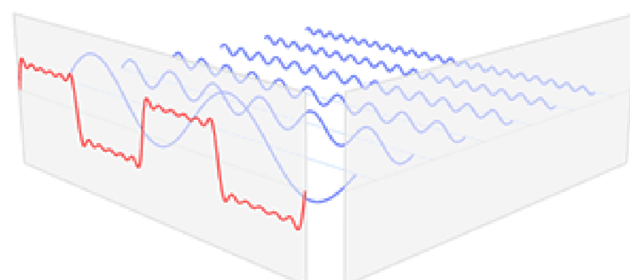

Figure 2. A visualization of a periodic function (red) decomposed into 6 cosine functions (blue) [30]

In our earlier paper [31], we demonstrated a similar model for sedentary behavior and found that some college students had clear daily and weekly sedentary behaviors such as being sedentary every Monday at 1PM as a result of class attendance. We consider this model "history-free" because once the active sedentary cycles have been discovered in model training phase, subjects' activity histories are not required to make predictions. What is needed are the trained model and time index $t$.

## 3.2. History-Dependent Sedentary Behavior Prediction (HD-SBP)

Another type of temporal behavior patterns is also commonly observed in nature. Behaviors with such temporal patterns are linearly correlated with the history of these behaviors themselves. For instance, after running a 400-meter race ($y_{t-1} = 20\%$, low sedentary level), for the next 5 minutes, a person is likely to sit and take a rest ($y_t = 80\%$, highly sedentary). In this example, the current state of sedentary behavior is negatively correlated with the previous state of the behavior. Mathematically, we can model this correlation as $y_t = \alpha\, y_{t-1} + 1$, where the coefficient $\alpha$ is negative ($-1$) in this case, indicating a negative correlation. Such linear correlation can be interpreted as a linear rhythm.

Linear correlation with previous behaviors may not always be as simple as first-order. Linear dependency can be long (Figure 3). To generalize our history-dependent model, we define it as:

$$Y_t = \sum_{i=1}^{n} \beta_i Y_{t-i} + \varepsilon \qquad (3)$$

where $n$ is the order of the model, $\varepsilon$ is a constant, and $\beta_i$ are the coefficients, which can be estimated with the Maximum Entropy Method (MEM) [32]. This type of history-dependent prediction has been discussed in our previous study [33] and showed good accuracy in terms of having low Mean Squared Errors (MSEs) in predicting sedentary behaviors.

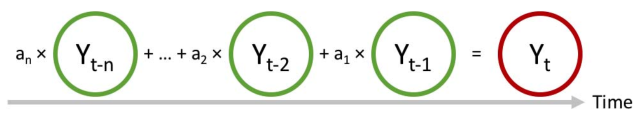


Figure 3. N-th Order Linear Correlation

We call this model "history-dependent" because it needs to remember and utilize the subject's history of sedentary behaviors $\{y_{t-1}, y_{t-2}, \ldots, y_{t-n}\}$ for the purpose of making predictions. It is instructive to note that the difference between above two predictive models (HF-SBP and HD-SBP) is the required input information—time index $t$ *vs.* historical states $\{Y_{t-1}, Y_{t-2}, \ldots, Y_{t-n}\}$.

## 3.3. Hybrid Sedentary Behavior Prediction (Hy-SBP)

To combine the benefits of the HF-SBP and HD-SBP models and synthesize a unified model, we propose a semi-history-dependent predictive model (Equation 4). This hybrid model (Hy-SBP) is a weighted combination of the HF-SBP and HD-SBP models. The HF-SBP part of this model captures the patterns caused by periodic cycles at multiple time scales and the HD-SBP part captures the patterns inside these cycles.

To build a Hy-SBP model, we first perform Fourier transform on the sequence of sedentary behaviors and find the dominant periodic functions—cosine functions with the largest amplitudes. These dominant periodic functions can be interpreted as the *cyclical rhythms* of sedentary behaviors. Then, we subtract cyclical rhythms from the sequence and get the remaining sedentary behaviors in the sequence as *linear rhythms*. We further model the linear rhythms with autoregression.

$$Y = \sum_{i=1}^{k} \alpha_i cos(\omega_i t + \phi_i) + \sum_{j=1}^{l} \beta_j \tilde{Y}_{t-j} + \varepsilon \qquad (4)$$

where $\varepsilon$ is a constant, $k$ is the number of cyclical rhythms, $l$ is the order of linear rhythm, $\omega_i$ are the frequencies of periodic functions, $\phi_i$ are the phase shifts, $\alpha_i$ are the amplitudes, $t$ is the time index, $\beta_j$ are the autoregressive coefficients, and $\tilde{Y_j}$ are the remaining sedentary behaviors in the sequence after subtracting cyclical rhythms.

An example is demonstrated in Figure 4. The algorithm for generating such Hy-SBP model is described with pseudocode in Algorithm 1.

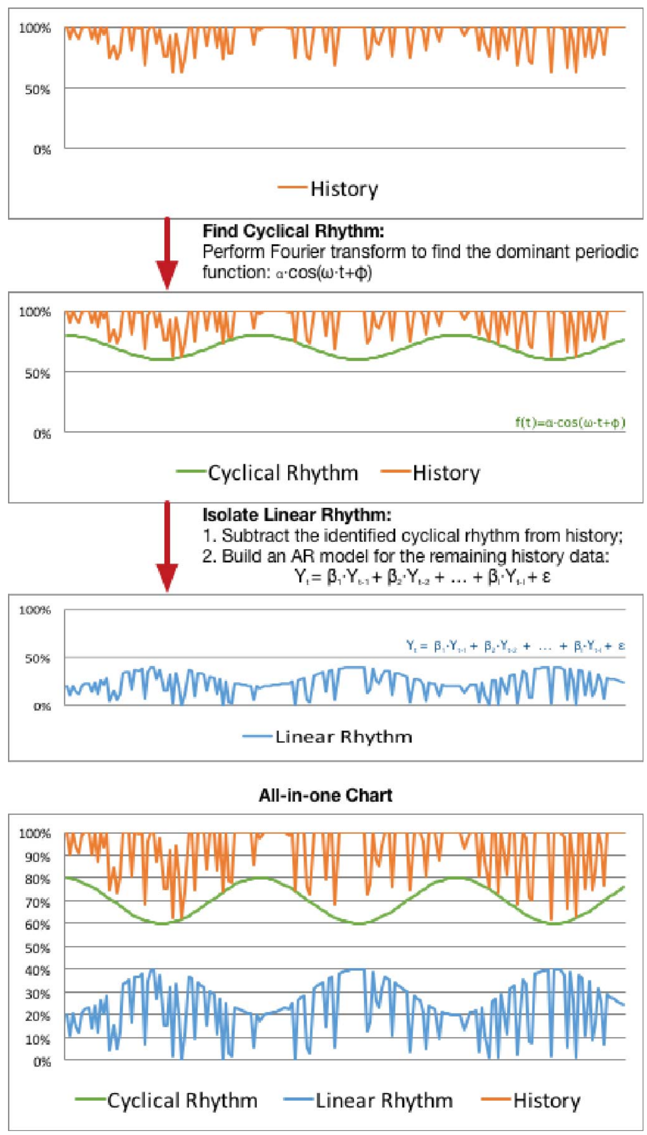


Figure 4. An example of generating Hy-SBP model

**Algorithm 1** Hy-SBP model generator: *genHybridModel(y, k, l)*

**Require:** $k \geq 0$, $l \geq 0$, and $y = [y_1, y_2, \ldots, y_t]$
**Ensure:** $k + l > 0$
$\omega, \phi, \alpha \leftarrow$ fourierTransform(y)
$\omega, \phi, \alpha \leftarrow \omega[1..k], \phi[1..k], \alpha[1..k]$
$\tilde{y} \leftarrow$ inverseFourierTransform($\omega, \phi, \alpha$)
$\hat{y} \leftarrow y - \tilde{y}$
$\beta, \varepsilon \leftarrow$ burgEstimateWithMaxEntropy($\hat{y}$, $l$)
**return** $\omega, \phi, \alpha, \beta, \varepsilon$

# 4. Experiment to Validate Our Proposed Predictive Models

In order to validate our proposed Hy-SBP model, we utilize data from a public dataset from Dartmouth College, called "StudentLife" [17]. StudentLife provides us with activity data (automatically sensed using smartphones) of a class of 49 Dartmouth students over a 10-week term (2013 Spring). The dataset contains students' activities, locations visited, and smartphone usage. The StudentLife dataset is publicly available on the project's website (http://studentlife.cs.dartmouth.edu). For our purposes, StudentLife dataset presents a complete academic term of time series of students' activity states, contextual information, smartphone usage, and physical activities, including sedentary behaviors that we utilize in validating our model.

The raw *physical activity* logs provided by StudentLife dataset contain timestamps and activity labels: *Stationary*, *Walking*, *Running*, and *Unknown*. The physical activities of subjects were sampled every 2–3 seconds in 1 of every 4 minutes. The activity types were inferred from acceleration data sensed with smartphone's accelerometer and classified using a decision tree classifier with 94% accuracy. In this study, we are only interested in whether the activity was *Stationary* (i.e. sedentary). We make the assumption that sedentary behaviors such as sitting and lying down probably correspond to the *Stationary* state in the StudentLife dataset. As such, non-sedentary behaviors such as standing still for 2–3 seconds may be mis-classified as *Stationary*, but is likely not be as common as sitting still for 2–3 seconds.

As defined in the previous section, ***sedentary behavior*** is defined as the percentage of activity logs classified as *Stationary* for a given person in a given *period of time*. In this experiment, we consider 1-hour buckets of time. For instance, between 08:00AM and 10:00AM, there are 2 time buckets: 08:00AM–09:00AM and 09:00AM–10:00AM. Sedentary behaviors occurring in a given period are summarized as a sedentary level—a percentage, 0–100%.

## 4.1. Estimate $k$ & $l$

To estimate parameters $k$ (the number of cyclical rhythms) and $l$ (the order of linear rhythm) in the Hy-SBP model, we exhaustively test different combinations of $k$ & $l$ in certain ranges and select the $\langle k, l \rangle$ pair with the smallest Mean Squared Error (MSE) tested on the training data (Algorithm 2). The MSE quantifies the difference between sedentary levels predicted by our model and the observed sedentary levels in the dataset. In this experiment, the length of historical data used for training is 2 weeks.

As mentioned in the previous section, the HF-SBP part of the Hy-SBP model captures the patterns across cycles while the HD-SBP part captures the patterns inside cycles. The $l$ parameter required in HD-SBP part is constrained to take values less than the length of the smallest cyclical rhythm. For instance, if the smallest cyclical rhythm is a half-day (12-hour) periodic pattern, $l$ should be $\leq 12$.

**Algorithm 2** Hy-SBP model optimizer: *genOptimalHybridModel(y, k, l)*

**Require:** $minK \leq maxK$, $minL \leq maxL$, and $y = [y_1, y_2, \ldots, y_t]$
**for** $k = minK$ to $maxK$ **do**
  **for** $l = minL$ to $maxL$ **do**
    model = genHybridModel(k, l, y)
    models[] $\leftarrow$ model
    mse[] $\leftarrow$ evaluate(model, y)
  **end for**
**end for**
i = findIndexOfMinimium(mse)
**return** models[i]

## 4.2. "Replay" and Model Students' Daily Lives

For each subject in the StudentLife dataset, we "replay" activity logs of their daily lives—feeding their historical sedentary behavior data (1-hour time buckets) in chronological order to our algorithm—and model them by building our predictive models. The algorithm moves along the sequence of buckets, building models, making predictions, and comparing predictions with the ground truth—the real sedentary behaviors in the data. At any point in the sequence of a subject's sedentary behaviors, we use the subject's past 2 weeks of sedentary behavior data (look back 336 1-hour buckets) preceding that point to generate an optimized Hy-SBP model, and then use the model to predict the subject's next sedentary behavior. This 2-week period acts like a sliding window.

For instance, *Subject 19* has 1,536 buckets sequenced chronologically. Initially, we use the first 336 buckets to generate a model, and use it to predict the 337th bucket. Then, we move the look-back window forward one step and buckets 2–337 to predict the 338th bucket. By the time all 1,536 buckets have been traversed for *Subject 19*, 1,200 predictions would have been made ($1536 - 336 = 1200$).

For all 49 subjects in StudentLife dataset, we built a total of 49,248 Hy-SBP models and made 49,248 predictions. The automatically optimized parameters ($k$ and $l$) used with Algorithm 2 are summarized in Figure 5, where the columns correspond to different values of $l$, the rows correspond to different values of $k$, and the numbers in the cells are the frequencies of the model selected for the corresponding parameter combinations.

*L*, the order of linear rhythm

*K*, the number of cyclical rhythms

| K/L | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | SUM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 23935 | 73 | 48 | 56 | 65 | 90 | 156 | 119 | 165 | 236 | 718 | 1724 | 27385 |
| 1 | 7519 | 57 | 17 | 40 | 36 | 20 | 50 | 55 | 108 | 69 | 161 | 385 | 8517 |
| 2 | 3219 | 19 | 5 | 10 | 10 | 3 | 21 | 25 | 48 | 47 | 95 | 440 | 3942 |
| 3 | 1523 | 12 | 3 | 9 | 5 | 11 | 24 | 19 | 56 | 32 | 69 | 352 | 2115 |
| 4 | 845 | 9 | 2 | 7 | 6 | 13 | 23 | 17 | 29 | 43 | 63 | 305 | 1362 |
| 5 | 537 | 2 | 2 | 7 | 6 | 6 | 15 | 16 | 48 | 42 | 36 | 286 | 1003 |
| 6 | 329 | 5 | 0 | 8 | 4 | 5 | 20 | 13 | 25 | 36 | 50 | 290 | 785 |
| 7 | 248 | 5 | 4 | 2 | 5 | 15 | 14 | 6 | 15 | 17 | 35 | 262 | 628 |
| 8 | 180 | 6 | 3 | 5 | 0 | 8 | 18 | 9 | 13 | 24 | 38 | 234 | 538 |
| 9 | 110 | 0 | 0 | 2 | 5 | 5 | 9 | 10 | 27 | 20 | 26 | 197 | 411 |
| 10 | 98 | 3 | 2 | 3 | 7 | 8 | 17 | 4 | 22 | 14 | 35 | 192 | 405 |
| 11 | 84 | 1 | 2 | 4 | 4 | 13 | 16 | 11 | 12 | 16 | 38 | 183 | 384 |
| 12 | 41 | 2 | 3 | 3 | 5 | 7 | 17 | 7 | 13 | 18 | 26 | 145 | 287 |
| 13 | 58 | 3 | 2 | 2 | 3 | 7 | 23 | 7 | 19 | 13 | 57 | 184 | 378 |
| 14 | 41 | 2 | 3 | 5 | 8 | 12 | 25 | 8 | 13 | 17 | 47 | 201 | 382 |
| 15 | 51 | 8 | 1 | 8 | 12 | 16 | 51 | 20 | 35 | 46 | 82 | 396 | 726 |
| SUM | 38818 | 207 | 97 | 171 | 181 | 239 | 499 | 346 | 648 | 690 | 1576 | 5776 | 49248 |

Figure 5. Frequency of Hy-SBP parameter combinations

## 5. Discussion

### 5.1. Cyclical Rhythms and Linear Rhythms

Setting the values of $k = 0$ and $l = 0$, makes the Hy-SBP model act like a simple Moving Average (MA) model—predicting the future with the simple average of the past. As shown in Figure 5, 23,935 times (48.6%) the *genOptimalHybridModel* algorithm (Algorithm 2) selected $k = 0$ and $l = 0$ as the optimal values to use in building the Hy-SBP model and to make predictions. Although this goes against our assumption that people's daily lives consist with cyclical rhythms ($k > 0$) and linear rhythms ($l > 0$), this is not surprising as people spend about 6–8 hours per day sleeping (sedentary level $= 100\%$). While a subject is asleep, the best prediction could be as simple as using the average sedentary level of the prior few hours, during which the subject may also be sleeping.

The first colored column in Figure 5 represents the Hy-SBP models with $l = 0$, which means the Hy-SBP models only have the HF-SBP part (for cyclical rhythms). The first colored row represents the Hy-SBP models with only the HD-SBP part (for linear rhythms). For this particular dataset, it is more common for Hy-SBP models to have only the cyclical (periodic) rhythms (14,883) than having only the linear (based on recent history) rhythms (3,450).

Multiple cyclical rhythms with different cycle lengths may occur simultaneously for each subject. Consider a subject who sits in a weekly class for 3 hours every Monday (weekly cycle) and also sits for lunch at the same time every day (daily cycle). Figure 6 shows how frequently different cycle lengths were recognized by the Hy-SBP models in the StudentLife dataset. Half-day rhythms, daily rhythms, weekly rhythms, and bi-weekly rhythms stood out. Daily rhythms were commonly observed. We speculate that such a strong pattern was largely caused by sleeping—a sedentary behavior albeit one that should not be intervened. We also suspect that the weekly rhythms were caused by scheduled sitting in classes—another sedentary behavior. In future work, by analyzing contextual information such as time and location, we will further investigate the causes of these cyclical rhythms. We make the point here that since our models are simple, it catches all sedentary behaviors including sedentary behaviors such as watching TV that are typically targeted by interventions, as well as sleeping and sitting at work or classes, which will likely not be the targets of interventions from a practical perspective.

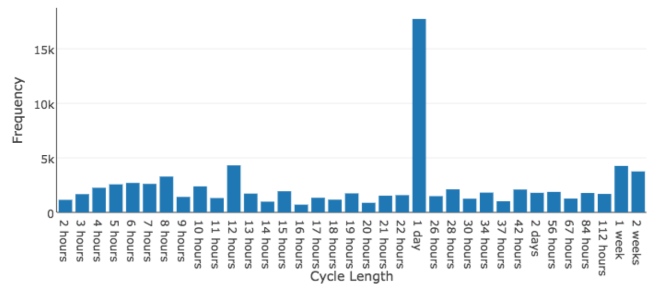


Figure 6. Frequencies of Hy-SBP models with different cyclical rhythms

Not all subjects had distinct cyclical and linear rhythms. For 13 of 49 subjects, our Hy-SBP model outperforms the baseline—moving average model—when predicting sedentary behaviors. For the rest, however, it does not always outperform the average model. We speculate that this might be caused by the diversity of patterns exhibited by students (the demographic covered by StudentLife dataset). Since students' schedules are generally more flexible than many other occupations, a wide variety of patterns (and models) are to be expected.

### 5.2. Model Sensitivity to the Length of Historical Data Used for Training

For the bulk of our data exploration experiments, we used a 2-week sliding window of historical data for training the optimized Hy-SBP models. In this section, we explore how sensitive our models are to the size of the look-back sliding window. For instance, if the sliding window is increased, does the accuracy of predicting sedentary behaviors improve? Intuitively, using more data to predict a person's behavior should yield a more accurate prediction. However, in statistical models, beyond a certain threshold, using more training data can cause the problem of overfitting [34].

To study how sensitive the accuracy of the Hy-SBP model is to the length of the sliding window, we compared the accuracies of Hy-SBP models built with different sliding window lengths for all students in the StudentLife dataset. Figure 7 shows the MSE of predictions—the smaller the MSE is, the more accurate the model is—made by the Hy-SBP model. As depicted in Figure 7, as the length of historical data increases, the average MSE of all students decreases.

However, increasing window size does not always yield more accurate predictions for every subjects (Figure 9). We further categorize the results as follows:

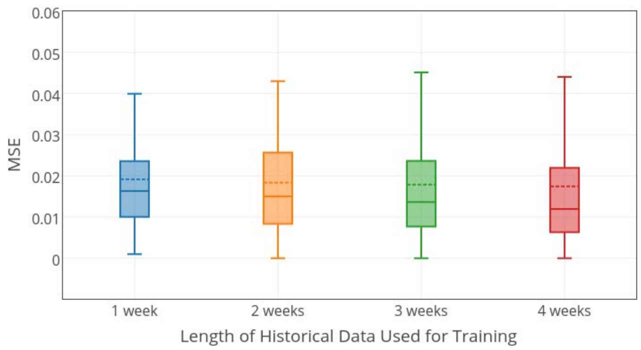


Figure 7. MSE vs. length of historical data used for training

1) ***Strictly obey***: Monotonically increasing. i.e. $MSE_{1week} > MSE_{2weeks} > MSE_{3weeks} > MSE_{4weeks}$;
2) ***Partially obey***: $MSE_{1week} > MSE_{4weeks}$ only but no clear pattern for 2-week and 3-week windows;
3) ***Strictly disobey***: Monotonically decreasing. i.e. $MSE_{1week} < MSE_{2weeks} < MSE_{3weeks} < MSE_{4weeks}$; and
4) ***Partially disobey***: $MSE_{1week} > MSE_{4weeks}$ only but no clear pattern for 2-week and 3-week windows.

Figure 8 summarizes the number of students in each of the above categories in the StudentLife dataset.

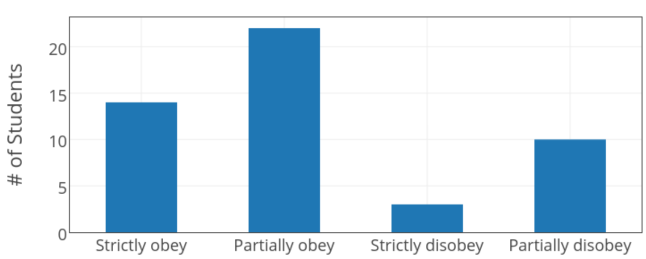


Figure 8. "More data, more accurate" pattern

## 5.3. Granularity of Time and Sedentary Level

Intuitively, predicting sedentary patterns in the near future (e.g., next 1 hour) is more difficult than forecasting longer term behaviors (e.g., next 6 hours), because people's daily lives are very complicated and dynamic. Shorter term predictions are more affected by random noise, which can be smoothed out in longer term predictions. Many factors may significantly influence peoples' next 1-hour sedentary behaviors. Making the bucket size bigger may increase the accuracy of prediction, but it will reduce the granularity of prediction.

In future, we will explore bucket sizes that establish a balance between prediction accuracy and granularity. We would like to investigate how far into the future (e.g., 1 bucket away or 2 buckets away) the Hy-SBP model can predict and how small the granularity of time we can predict (e.g., next 10 minutes or next 20 minutes). For sedentary behavior interventions, we believe that sub-hour predictions may be more useful. This requires further investigation and user studies to confirm.

In this study, our original goal was bold—we wanted to predict the exact future sedentary level (e.g., 84%) of a subject. An easier task may be predicting discretized ranges of sedentary levels, e.g., *very sedentary* (90-100%), *sedentary* (80-90%), *active* (40-80%), and *very active* (0-40%). We will explore such discretization in the future.

## 5.4. Model Complexity

The context-free model discussed in this paper is a light-weight predictive model, which only models the patterns of sedentary behavior itself. No contextual information (e.g., location and temperature) is exploited to improve the accuracy of prediction. It has two advantages over the context-aware models: fewer sensors required and lower model complexity—the complexity of the model decreases as the number of required context inputs decrease [35].

For wearable devices such as Fitbit, a light-weight predictive model will be more computationally feasible than models that utilize large multivariate contextual information. Resource-constrained mobile/wearable devices might not be equipped with the sensors required to sense rich context and might not have sufficient computational power to discover patterns and to make predictions. For more powerful devices such as a smartphone, the cost of gathering and analyzing various types of contextual information to generate sophisticated context-aware models will be more reasonable.

While the models discussed in this study do not take advantage of context information for prediction purpose, we by no means underestimate the usefulness of context. In our previous work [36], we found that using a simple Logistic Regression model and smartphone-sensed context information such as user location, time, and smartphone app usage, we could predict whether a student will be very sedentary in the next hour with a precision of 73.1% (recall of 87.7%).

In future, we would like to research the idea of combining our context-free models (discussed in this paper) and the context-aware models (discussed in [36]) and to develop a unified framework for model evaluation.

# 6. Conclusion

In this paper, we proposed two canonical types of model and a hybrid model to quantitatively perform *rhythm analysis*, a concept proposed by Lefebvre [25]. Cyclical rhythms and linear rhythms, proposed by Lefebvre, are modeled using periodic functions (history-free model) and linear functions (history-dependent model) respectively. Our preliminary results with our Hy-SBP model showed that cyclical rhythms are more common than linear rhythms, and

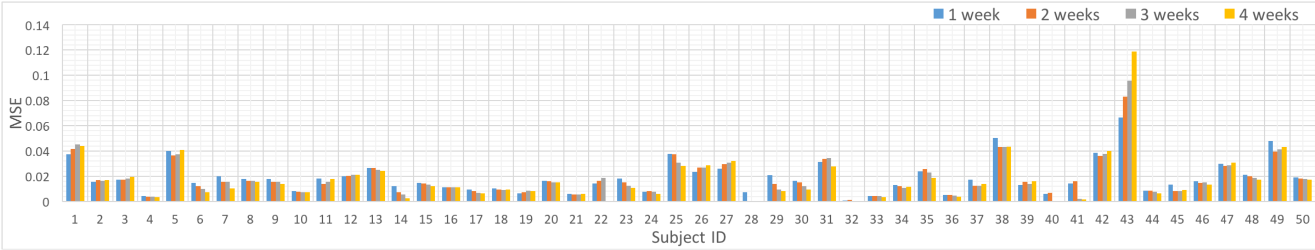


Figure 9. MSE when Different Lengths of Historical Data are Used for Training the Hy-SBP Model

half-day rhythms, daily rhythms, weekly rhythms, and bi-weekly rhythms were observed clearly among students in the StudentLife dataset.

In future, we would like to see how *Just-In-Time* health behavior intervention can be integrated with behavior prediction to achieve better health behavior change outcomes. For instance, if it is predicted that a subject will become sedentary (e.g., watching TV) after dinner, a reminder might be sent to her/his smartphone just as s/he arrives home from work, reminding her/him to walk around during the TV commercial breaks.

Finally, we believe that beyond sedentary behaviors, the concept of rhythm analysis could apply broadly to many health-related behaviors. We believe that it could be applicable to biological and social rhythms and might help mitigate chronic behavioral disorders (e.g., unhealthy dietary habits, cigarette smoking, and substance abuse) [14] and mental disorders (e.g., cyclothymic and bipolar disorders) [15]. Our proposed framework may help facilitate various interventions driven by behavior prediction, helping people live healthier lives.

## Acknowledgments

We sincerely thank the StudentLife team for sharing their data publicly online, making it possible for other researchers to make more discoveries from their dataset [17].

## References

[1] R. R. Pate, J. R. O'Neill, and F. Lobelo, "The evolving definition of "sedentary"." *Exercise and sport sciences reviews*, vol. 36, no. 4, pp. 173–8, 10 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18815485

[2] E. G. Wilmot, C. L. Edwardson, F. A. Achana, M. J. Davies, T. Gorely, L. J. Gray, K. Khunti, T. Yates, and S. J. H. Biddle, "Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis," *Diabetologia*, vol. 55, no. 11, pp. 2895–2905, 2012. [Online]. Available: http://dx.doi.org/10.1007/s00125-012-2677-z

[3] E. Stamatakis, M. Hamer, and D. W. Dunstan, "Screen-based entertainment time, all-cause mortality, and cardiovascular events: population-based study with ongoing mortality and hospital events follow-up." *Journal of the American College of Cardiology*, vol. 57, no. 3, pp. 292–9, 1 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0735109710044657

[4] N. Owen, T. Sugiyama, E. E. Eakin, P. A. Gardiner, M. S. Tremblay, and J. F. Sallis, "Adults' sedentary behavior determinants and interventions." *American journal of preventive medicine*, vol. 41, no. 2, pp. 189–96, 8 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0749379711003229

[5] J. Buckworth and C. Nigg, "Physical activity, exercise, and sedentary behavior in college students." *Journal of American college health : J of ACH*, vol. 53, no. 1, pp. 28–34, 1 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.3200/JACH.53.1.28-34

[6] C. Fountaine, G. Liguori, A. Mozumdar, and J. Schuna, "Physical Activity and Screen Time Sedentary Behaviors in College Students," 2011. [Online]. Available: http://digitalcommons.wku.edu/ijes/vol4/iss2/3

[7] S. L. Gortmaker, "Television Viewing as a Cause of Increasing Obesity Among Children in the United States, 1986-1990," *Archives of Pediatrics & Adolescent Medicine*, vol. 150, no. 4, p. 356, 4 1996. [Online]. Available: http://archpedi.jamanetwork.com/article.aspx?articleid=517896

[8] M. H. Proctor, L. L. Moore, D. Gao, L. A. Cupples, M. L. Bradlee, M. Y. Hood, and R. C. Ellison, "Television viewing and change in body fat from preschool to early adolescence: The Framingham Children's Study." *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity*, vol. 27, no. 7, pp. 827–33, 7 2003. [Online]. Available: http://dx.doi.org/10.1038/sj.ijo.0802294

[9] K. C. Herman, K. W. Merrell, W. M. Reinke, and C. M. Tucker, "The role of school psychology in preventing depression," *Psychology in the Schools*, vol. 41, no. 7, pp. 763–775, 9 2004. [Online]. Available: http://doi.wiley.com/10.1002/pits.20016

[10] H. Lefebvre, *Critique of Everyday Life*. Verso, 1991. [Online]. Available: https://books.google.com/books?id=5FzpoAEACAAJ

[11] L. S. Noelker and R. Browdie, "Sidney Katz, MD: A New Paradigm for Chronic Illness and Long-Term Care," *The Gerontologist*, vol. 54, no. 1, pp. 13–20, 2 2014. [Online]. Available: http://gerontologist.oxfordjournals.org/cgi/doi/10.1093/geront/gnt086

[12] O. Rantala and A. Valtonen, "A rhythmanalysis of touristic sleep in nature," *Annals of Tourism Research*, vol. 47, pp. 18–30, 7 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0160738314000371

[13] J. McQuoid, A. Griffin, S. S. J. McQuoid, and S. Sharpe, "Rhythm & Habit: Patient Self-Management of Chronic Disease."

[14] B. P. Hasler, L. J. Smith, J. C. Cousins, and R. R. Bootzin, "Circadian rhythms, sleep, and substance abuse," *Sleep Medicine Reviews*, vol. 16, no. 1, pp. 67–81, feb 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1087079211000335

[15] C. A. McClung, "Circadian genes, rhythms and the biology of mood disorders," *Pharmacology & Therapeutics*, vol. 114, no. 2, pp. 222–232, may 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0163725807000368

[16] Google Inc., "Android APIs - Google Fit." [Online]. Available: https://developers.google.com/fit/android/

[17] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*. New York, New York, USA: ACM Press, 9 2014, pp. 3–14. [Online]. Available: http://dl.acm.org/citation.cfm?id=2632048.2632054

[18] E. M. Tapia, S. S. Intille, and K. Larson, "Activity Recognition in the Home Using Simple and Ubiquitous Sensors," 2004, pp. 158–175. [Online]. Available: http://link.springer.com/10.1007/978-3-540-24646-6_10

[19] Y. Fu, Ed., *Human Activity Recognition and Prediction*. Cham: Springer International Publishing, 2016. [Online]. Available: http://link.springer.com/10.1007/978-3-319-27004-3

[20] A. Pentland and A. Liu, "Modeling and Prediction of Human Behavior," *Neural Computation*, vol. 11, no. 1, pp. 229–242, 1 1999. [Online]. Available: http://www.mitpressjournals.org/doi/abs/10.1162/089976699300016890

[21] L. E. Baum and T. Petrie, "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 12 1966. [Online]. Available: http://projecteuclid.org/euclid.aoms/1177699147

[22] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 282–289. [Online]. Available: http://dl.acm.org/citation.cfm?id=645530.655813

[23] D. J. Cook and N. C. Krishnan, *Activity Learning, Discovering, Recognizing, and Predicting Human Behavior from Sensor Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2 2015. [Online]. Available: http://doi.wiley.com/10.1002/9781119010258

[24] A. Galata, N. Johnson, and D. Hogg, "Learning Variable-Length Markov Models of Behavior," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 398–413, 3 2001. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1077314200908944

[25] H. Lefebvre, G. Moore, and S. Elden, *Rhythmanalysis: Space, Time and Everyday Life*, ser. Bloomsbury Revelations. Bloomsbury Publishing, 2013. [Online]. Available: https://books.google.com/books?id=ToDjBAAAQBAJ

[26] T. Heath, T. Oc, and S. Tiesdell, *Public Places - Urban Spaces*. Taylor & Francis, 2011. [Online]. Available: https://books.google.com/books?id=EYtNvjLVisUC

[27] J. McQuoid, J. Welsh, L. Strazdins, A. L. Griffin, and C. Banwell, "Integrating paid work and chronic illness in daily life: A space-time approach to understanding the challenges," *Health & Place*, vol. 34, pp. 83–91, 7 2015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S135382921500060X

[28] B. Nansen, "Step-counting: The anatomo- and chrono-politics of pedometrics," *Continuum: Journal of Media & Cultural Studies*, vol. 22, no. 6, pp. 793–803, 12 2008. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/10304310802452453

[29] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–297, 5 1965. [Online]. Available: http://www.ams.org/jourcgi/jour-getitem?pii=S0025-5718-1965-0178586-1

[30] L. Barbosa, "Fourier transform time and frequency domains," 2013. [Online]. Available: https://commons.wikimedia.org/wiki/File:Fourier_transform_time_and_frequency_domains_(small).gif

[31] Q. He and E. Agu, "A frequency domain algorithm to identify recurrent sedentary behaviors from activity time-series data," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2016, pp. 45–48.

[32] J. P. Burg, "Maximum Entropy Spectral Analysis," Ph.D. dissertation, Stanford University, 1975. [Online]. Available: http://sepwww.stanford.edu/data/media/public/oldreports/sep06/

[33] Q. He and E. Agu, "Towards Sedentary Lifestyle Prevention: An Autoregressive Model for Predicting Sedentary Behaviors," in *IEEE 10th International Symposium on Medical Information and Communication Technology (ISMICT '16)*, Worcester, MA, USA, 2016.

[34] D. M. Hawkins, "The Problem of Overfitting," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 1, pp. 1–12, 1 2004. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/ci0342472

[35] K. Henricksen and J. Indulska, "Developing context-aware pervasive computing applications: Models and approach," *Pervasive and Mobile Computing*, vol. 2, no. 1, pp. 37–64, feb 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1574119205000441

[36] Q. He and E. O. Agu, "Smartphone usage contexts and sensable patterns as predictors of future sedentary behaviors," in *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*. IEEE, nov 2016, pp. 54–57. [Online]. Available: http://ieeexplore.ieee.org/document/7797695/