Automated Selection of Changepoints using Empirical P-values and Trimming (ASCEPT)

Matthew Quinn[1], Arlene Chung[2-4], Kimberly Glass[1,5-6]*

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[2]School of Information & Library Science & Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA [3]Division of General Medicine & Clinical Epidemiology, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, USA [4]Google Health, Mountain View, CA, USA, [5]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA, [6]Department of Medicine, Harvard Medical School, Boston, MA, USA

*Corresponding author: kimberly.glass@channing.harvard.edu
181 Longwood Ave., Boston, MA 02115
(617) 525-2715

Keywords: mobile health, changepoints, time series, monte carlo method, regression

Word count, excluding title page, abstract, lay summary, funding, author contributions, conflict of interest, data availability, references, figures, and tables: 3,983

ABSTRACT

Objectives

One challenge when analyzing mobile health (mHealth) data are patterns that result from updates to the proprietary algorithms that process these data. Since the timing of these updates are not publicized, an analytic approach is necessary to determine whether changes in mHealth data are due to lifestyle behaviors or algorithmic updates. Existing methods for identifying changepoints do not consider multiple types of changepoints, may require pre-specifying the number of changepoints, and often involve non-intuitive parameters, such as an optimization penalty. We propose a novel approach, Automated Selection of Changepoints using Empirical P-values and Trimming (ASCEPT), to select an optimal set of changepoints in mHealth data.

Materials and Methods

ASCEPT involves two stages: (1) identification of a statistically significant set of changepoints from sequential iterations of a changepoint detection algorithm; and (2) trimming changepoints within linear and seasonal trends. ASCEPT is available at https://github.com/matthewquinn1/changepointSelect.

Results

We demonstrate ASCEPT's utility using real-world mHealth data collected through the Precision VISSTA study. We also demonstrate that ASCEPT outperforms a comparable method, Circular Binary Segmentation (CBS), and illustrate the impact when adjusting for changepoints in downstream analysis.

Discussion

ASCEPT offers a practical approach for identifying changepoints in mHealth data that result from algorithmic updates. ASCEPT's only required parameters are a significance level and goodness-of-fit threshold, offering a more intuitive alternative compared to other approaches.

Conclusion

ASCEPT provides an intuitive and useful way to identify which changepoints in mHealth data are likely the result of updates to the underlying algorithms that process the data.

LAY SUMMARY

Mobile health (mHealth) has taken on an increasingly important role in medicine and public health. However, how the mHealth data are reported to an end user can change based on updates to the underlying proprietary algorithms that process these data. The times at which these changes occur represent potential "changepoints". Effectively using mHealth data for health applications will require correctly identifying when such algorithmic changes occur and distinguishing them from changes in mHealth data that are due to lifestyle behaviors. Here we present Automated Selection of Changepoints using Empirical P-values and Trimming (ASCEPT), a method to identify changepoints in mHealth data. ASCEPT correctly identifies algorithmic changepoints both in the context of simulated data and real Fitbit data collected as part of the Precision VISSTA study. We also find that it outperforms a comparable approach, Circular Binary Segmentation (CBS), and that this difference in performance has a clear impact

when adjusting for the identified changepoints. ASCEPT offers an intuitive and useful approach for identifying changepoints in mHealth data before performing any downstream analysis.

INTRODUCTION

In recent years, mobile health (mHealth) has taken on a growing importance in medicine and public health, among other fields.[1–3] mHealth devices often produce time series by recording variables, such as heart rate and number of steps, at regular intervals (e.g., hourly or daily). Studying these data can bring important insights into how health changes over time. For instance, an individual might greatly reduce their daily number of steps upon becoming ill. This type of event is associated with a "changepoint", a time at which the distribution of data changes, and typically corresponds to a change in the mean of the data, or a "mean-shift". However, in addition to changepoints due to lifestyle or behavioral changes, wearable mHealth devices also have both planned updates and unexpected technical issues that can impact data collection and reporting. These can introduce "technological changepoints" to the data, which can be difficult to distinguish from behaviorally driven changes, obscuring patterns of interest. It is therefore necessary to identify and correct for these technological changepoints before proceeding with downstream analysis.

Unfortunately, the timing of planned updates and identified technological issues is neither systematically documented nor generally publicized by mHealth device manufacturers. Although an investigator could potentially monitor a manufacturer's release notes to determine when updates are pushed or manually inspect the mHealth data to find potential technological changepoints, these approaches are neither scalable nor practical, and are especially challenging when data collection comes from multiple types of devices. Even within a single manufacturer, updates are not necessarily pushed to all devices simultaneously, and some may require the user to first update an associated smartphone app.

There are several existing approaches that detect changepoints in time series by solving an optimization problem,[4] including Pruned Exact Linear Time (PELT).[5] Using PELT generally entails specifying an optimization penalty when detecting multiple changepoints, which is difficult to do in practice. Changepoints for a Range of PenaltieS (CROPS)[6] allows one to efficiently run PELT under various penalties, but does not select a final or optimal set of changepoints. Thus, instead of proposing another method for changepoint "detection", in this paper we develop Automated Selection of Changepoints using Empirical P-values and Trimming (ASCEPT), to identify changepoints in mHealth data through changepoint "selection". ASCEPT performs multiple runs of PELT, considering iteratively larger sets of changepoints until the selected set would no longer offer a statistically significant improvement over the prior set. Next, ASCEPT removes changepoints that are likely to be associated with lifestyle or behavioral changes rather than technological issues, ultimately yielding a single optimal set of changepoints. It is worth noting that ASCEPT also shares similarities with Circular Binary Segmentation (CBS),[7] which performs "pruning" to identify a subset of statistically significant changepoints. However, CBS does not consider features common to mHealth data, such as seasonal patterns. For a more detailed review of changepoint detection, please refer the **Supplementary Material**.

In this manuscript, we present results from running ASCEPT on both simulated data and Fitbit mHealth data collected by the Precision VISSTA study.[8] We find that ASCEPT appropriately identifies changepoints in the simulated data and most of the Precision VISSTA variables. We also compare the performance of ASCEPT to that of CBS[7] and show that ASCEPT provides better changepoint selection under comparable settings. We further demonstrate that these differences can have a clear impact when correcting the mHealth data based on the identified changepoints. Overall, ASCEPT provides an effective manner to objectively identify technological changepoints before continuing with downstream analysis.

MATERIALS AND METHODS

Data

Precision VISSTA Data

We evaluated the performance of ASCEPT on mHealth data collected by the Precision VISSTA study.[8] This study's data set includes users of many different devices. Due to their prevalence, we chose to focus on individuals who used a heart rate (HR) Fitbit device introduced in 2016-2019 (i.e., the Alta HR, Blaze, Charge 2, Charge 3, Inspire HR, Ionic, Versa, or Versa 2), multiple Fitbit devices, or an unknown Fitbit device. This subset of the data includes 203,351 observations on 298 individuals recorded between May 15th, 2015 and October 27, 2019.

These data include 7 activity variables (steps, distance, floors, elevation, calories, and time active) and 5 sleep variables (total sleep, deep sleep, light sleep, REM sleep, time awake at night, and times woken). The median number of users contributing data on a given day ranged from 50 for REM, to 93-95 for the other sleep variables, to 131 for the activity variables. In our analysis, we excluded floors and elevation as their values largely stay within narrow ranges near zero over the study period. We also excluded REM sleep due to a lack of any data between May 20, 2016 and March 26, 2017.

To help identify population-level changepoints, we focused on studying the daily median value of each variable across all users. **Figure 1A** shows the daily median amount of deep sleep, which experiences an abrupt shift around July 2017. While it is possible that a single individual might suddenly experience large changes in deep sleep due to various life events, like an injury or the birth of a child, it is unlikely that the median deep sleep across many users would truly decrease by 5-6 hours after July 19, 2017, only for it to later rebound multiple times. Instead, these shifts are likely due to changes in how Fitbit calculated deep sleep. It is critical to identify and control for these technological changepoints in order to correctly describe human behavioral changes relevant to health and disease.

Simulated Data

When shifts and patterns appear in real data, they are often not defined well enough to serve as a gold standard. In **Figure 1A**, there appears to be seasonality prior to July 19, 2017, but it is inconsistent. Likewise, it is challenging to determine whether some points between May 15, 2015 and May 15, 2016 constitute behaviorally driven or technological changepoints because the data only reflect contributions from between 7 and 54 unique users each day. Therefore, large behaviorally driven fluctuations are more likely during this period compared to later, when up to 160 unique users contributed sleep observations (**Supplementary Figure 1**).

Due to these limitations, we first demonstrate ASCEPT using a simulated time series containing 800 observations (**Figure 1B**). These data have sudden mean-shift changepoints at indices 49, 60, 600, 699, and 700, an increasing linear trend between indices 201 and 400 inclusive, and a seasonal pattern between indices 401 and 600 inclusive.

ASCEPT Stage 1: Changepoint Selection using Empirical P-values

The first stage of ASCEPT incrementally includes more mean-shift changepoints detected by PELT[5] until the newly proposed changepoints do not offer a statistically significant improvement in goodness-of-fit.

To begin, we let a changepoint at position j indicate that the time series' distribution changes between j and $j+1$. $\mathcal{T}_k$ will denote the set of changepoints detected by step k, where $\mathcal{T}_0 = \emptyset$, such that no changepoints are initially detected. This corresponds with imposing a large optimization penalty with PELT. From step k, CROPS[6] decreases the optimization penalty associated with PELT to find the next set of changepoints, denoted as $\mathcal{T}_{k+1}^*$. $\mathcal{T}_k$ will normally, but not necessarily, be a subset of $\mathcal{T}_{k+1}^*$. **Figures 2A** and **2B** depict a scenario in which we have detected changepoints $\mathcal{T}_k = \{305, 600\}$ and are evaluating $\mathcal{T}_{k+1}^* = \{49, 60, 305, 600\}$ as providing a significant improvement.

To assess whether or not $\mathcal{T}_{k+1}^*$ offers a significant improvement in goodness-of-fit, we must both choose a goodness-of-fit measure and assess its null distribution. For goodness-of-fit, we use the log-likelihood of normally distributed data. More specifically, between any two changepoints, or between a changepoint and the start or end of the series, the observations form a “segment”. We assume that all observations are independent and normally distributed, but that those within the same segment are also identically distributed. This assumption largely follows the implementation of PELT in R's “changepoint” package.[9]

We next assess the null hypothesis that $\mathcal{T}_k$ represents all of the true mean-shift changepoints in the time series. We do this is a manner that does not rely on asymptotic results, since mHealth time series can contain very small segments. We first generate a time series under the null by randomly drawing from normal distributions with the same means and standard deviations as the corresponding segments in the observed data. For example, **Figure 2A** shows the simulated data split into three segments by two changepoints at indices 305 and 600. **Figure 2C** illustrates a corresponding random sampling from the normal distributions that best fit each of these three segments. We record the log-likelihood for this random time series using the $\mathcal{T}_k$ changepoints. We then impose the changepoints in $\mathcal{T}_{k+1}^*$ onto this random time series and calculate the corresponding log-likelihood, as depicted in **Figure 2D**. Finally, we record the change in the log-likelihood under the null, comparing $\mathcal{T}_{k+1}^*$ with $\mathcal{T}_k$.

We repeat this process N times in order to calculate an empirical p-value for the observed change in the log-likelihood. If the observed change is statistically significant at the chosen level, α, then we reject the null that $\mathcal{T}_k$ represents all the true mean-shift changepoints for the time series, and instead select $\mathcal{T}_{k+1}^*$ as the current set of changepoints, $\mathcal{T}_{k+1}$. The procedure continues, comparing $\mathcal{T}_{k+1}$ to $\mathcal{T}_{k+2}^*$ and so forth, until we obtain a statistically insignificant result, as depicted in **Figure 1B**. This hypothesis testing process is called a “fixed-sequence” procedure and controls the family-wise error rate at the chosen significance level, α.[10]

ASCEPT Stage 2: Trimming Changepoints within Linear or Seasonal Trends

The changepoints identified in Stage 1 will include both technological changepoints, such as those associated with manufacturer updates, and changepoints from behaviorally driven patterns. In order to distinguish the former from the latter we note that, while software updates are likely to induce sudden mean-shifts in population-level mHealth data, behaviorally driven changes are more likely to be associated with linear or seasonal trends (see, for example, **Figure 1A**). For instance, individuals may walk more during the summer than the winter, or at the end of an exercise program compared to at the start. These trends technically contain a mean-shift at each point, but since these are all part of the same behaviorally driven pattern, ASCEPT aims to identify and remove them; for convenience we refer to these as “nuisance changepoints”. In contrast, changepoints that correspond with a sudden mean-shift, or that are at the start or end of a linear or seasonal trend, are considered “relevant changepoints” and retained (**Figure 1C**). We will refer to ASCEPT’s process of removing nuisance changepoints as “trimming”. Although it is the same principle as “pruning” used by methods such as CBS,[7] we avoid the term “prune” since it is also used to describe part of PELT's optimization process.[5]

We illustrate ASCEPT’s trimming process in **Figure 3**. **Figure 3A** shows a set of true changepoints identified by Stage 1. For every changepoint, we perform two types of model fits. We first fit piecewise linear and harmonic regressions on each of the two segments located to either side of the changepoint. We then fit linear and harmonic regressions across the two segments, ignoring the changepoint. To fit the linear models, we regress the values in a segment against their indices. For harmonic regressions, we first estimate a segment's period using the frequency associated with the peak of the periodogram and then fit the harmonic regression with a linear model based on this estimate. For each type of model fit, we calculate the root mean square error (RMSE). For relevant changepoints, the piecewise fits should greatly outperform the cross-segment fits. However, for nuisance changepoints that are part of an ongoing linear or seasonal trend, the best cross-segment and piecewise fits will perform similarly.

To illustrate this, **Figure 3B** shows a sudden mean-shift at index 60. Here, the best piecewise fit outperforms the best cross-segment fit by nearly a factor of three, suggesting that this is a relevant changepoint. In contrast, **Figure 3C** shows a nuisance changepoint that is within a linear trend. In this case, a linear regression across both segments performs only marginally worse than the best piecewise fit to the segments. Similarly, **Figure 3D** shows a changepoint within a seasonal pattern. Here, the cross-segment harmonic regression performs only marginally worse than the best piecewise fit.

This process of fitting piecewise and cross-segment models is done for every changepoint identified in Stage 1. For each changepoint, we record the ratio of the RMSE for the best cross-segment fit to the RMSE for the best piecewise fit. The changepoint that corresponds to the smallest ratio is then removed if it falls below a chosen “trimming threshold”. This process repeats for the remaining changepoints until no ratio falls below the threshold, as depicted in **Figure 1B**.

Segment Correction

We used changepoints identified by ASCEPT and CBS and fit constant, linear, and harmonic regressions to the corresponding segments. A linear or harmonic regression was deemed the best fit to a segment if the ratio of the constant fit's RMSE to the best corresponding linear or harmonic regression's RMSE was greater than a given “fitting threshold”. When the best fit was a linear or harmonic regression, we de-trended or de-seasonalized that segment. We then shifted and scaled all segments to match the location and scale of a chosen reference segment. The location was defined as the mean of the reference segment before any correction was performed and the scale was defined as the residual standard error for the best fitting model on that segment.

Parameters

For all main text analyses, Stage 1 of ASCEPT was run using a significance level of $\alpha = 0.01$ and $N = 10{,}000$ Monte Carlo simulations. Stage 2 was run using a trimming threshold of 1.2. That is, changepoints whose best cross-segment fit had an RMSE within 20% of the best piecewise fit were subject to being removed. We show results from running ASCEPT on the simulated data using various trimming thresholds in **Supplementary Figure 2**.

When running CBS, as implemented in R’s “DNAcopy” package,[11] we used a significance level of $\alpha = 0.01$ and 10,000 permutations. CBS's pruning threshold was set to 0.5; this yielded comparable results to ASCEPT’s 1.2 trimming threshold in terms of the number of changepoints identified per time series.

For segment correction, we used a fitting threshold of 1.75. Shifting and scaling were performed with respect to the seasonal segment, as captured by either ASCEPT or CBS, as the reference.

R Package

ASCEPT is implemented in “changepointSelect”, an R package hosted on GitHub at https://github.com/matthewquinn1/changepointSelect.

RESULTS

ASCEPT on the Simulated Data

We first applied ASCEPT to simulated data. We found that Stage 1 detected relevant changepoints at indices 49, 60, 225, 400, 600, 699, and 700, as well as many nuisance changepoints that were subsequently trimmed in Stage 2, as shown in **Figure 4A**. These results indicate that immediately after indices 49, 60, 225, 400, 600, 699, and 700, the simulated data experienced a statistically significant mean-shift that is not attributable to an ongoing linear or seasonal trend. Among the relevant changepoints, five correspond to sudden mean-shifts, while the other two segment off the linear and seasonal trends.

ASCEPT on the Precision VISSTA Data

Next, we applied ASCEPT to mHealth data from the Precision VISSTA study. **Figures 4B** and **4C** show the results for deep and light sleep. We observe similar results for these variables, which was expected because both contribute to total sleep. For deep sleep, ASCEPT identified changepoints on July 19th, 2017, September 1st, 2017, September 6th, 2017, February 14th, 2018, and February 15th, 2018. For light sleep, it identified changepoints on July 19th, 2017, August 9th, 2017, August 31st, 2017, September 6th, 2017, February 14th, 2018, and February 15th, 2018. Based on this analysis, we hypothesize that these are the dates immediately after which Fitbit changed how it calculated sleep stage information, impacting the relationship between deep and light sleep.

We further assessed these changepoints by cross-referencing with online information and found that some of the identified changepoints correspond to known firmware updates and glitches. Alta HR received firmware update 26.62.6 between August 1st, 2017 and August 10th, 2017,[12] corresponding to the August 9th, 2017 changepoint for light sleep. Likewise, Fitbit overhauled its calculation of sleep by introducing “Sleep Stages”, starting on March 6th, 2017.[13] Glitches with Sleep Stages were reported from within a week of the release through July 24th, 2017 for Alta HR, Blaze, and Charge 2 devices,[14] encompassing the changepoint on July 19, 2017. Blaze was also subject to glitches between September 3rd, 2017 and September 7th, 2017,[15] corresponding to the September 6th, 2017 changepoint.

Next, we used the daily median total sleep as a negative control (**Figure 4D**). While there were some large fluctuations in median total sleep during 2015, this variance is likely due to fewer individuals (as little as 7; see **Supplementary Figure 1**) contributing data on any given day. Accordingly, ASCEPT did not identify any changepoints in this time series after trimming. We do not suspect that Fitbit changed the calculation of total sleep during the period of the study.

Comparison between ASCEPT and CBS

ASCEPT shares some principles with other methods, such as CBS[7] (see **Supplementary Material**). Therefore, we compared the changepoints identified by CBS to those identified by ASCEPT. For the simulated data (**Figure 5A**), we found that CBS failed to capture the single-point segment at index 700, while ASCEPT successfully did. ASCEPT also successfully segmented off the linear and seasonal trends, while CBS split the linear trend into four segments.

We also compared ASCEPT and CBS on mHealth data from the Precision VISSTA study. For most variables, the two procedures yielded similar changepoints, although there were some important differences. For example, CBS failed to detect changepoints for the single-day shift in deep sleep on February 15th, 2018, while ASCEPT did (**Figure 5B**). The two procedures also differed greatly when applied to the times woken variable (**Figure 5C**). In particular, CBS failed to capture multiple changepoints from late 2017 to early 2018 and did not trim two nuisance changepoints that appear to be within linear or seasonal trends. In contrast, ASCEPT successfully captured the major relevant changepoints and trimmed nuisance changepoints. Comparisons of ASCEPT and CBS for the remaining mHealth variables are provided in **Supplementary Figures 3-4** and demonstrate that ASCEPT generally outperforms CBS when applied to real-world data.

While ASCEPT's primary purpose is to select changepoints, we also performed a simple correction to demonstrate the importance of accurately identifying changepoints (see **Methods**). In particular, we found the best fit model for each segment (**Figures 6A-B**) and then adjusted the data to match the location and scale of the segment containing the seasonal pattern, which was accurately identified by ASCEPT as indices 401 to 600 inclusive and identified by CBS as indices 356 to 600 inclusive. If the changepoints were accurately identified, then we expect the transformed time series to look like normally distributed noise without any mean-shifts. We found this to be true for the ASCEPT segment-corrected time series (**Figure 6C**). In contrast, the

CBS segment-corrected time series (**Figure 6D**) still contains trends, seasonality, and other mean-shifts due to the less accurate identification of changepoints. Results when using other fitting thresholds are shown in **Supplementary Figures 5-6**.

DISCUSSION AND CONCLUSIONS

We have developed an approach, ASCEPT, for identifying changepoints in mHealth data. ASCEPT builds upon the current state-of-the-art method, PELT, by incorporating the principles of statistical significance and trimming. In particular, ASCEPT adopts progressively larger sets of changepoints until the newly proposed set does not provide a statistically significant improvement in goodness-of-fit. ASCEPT then trims changepoints within linear or seasonal trends; in mHealth data, these changepoints are often the result of behavioral or lifestyle changes rather than technological issues. This results in a set of changepoints that can be used to adjust mHealth data prior to additional downstream analysis.

ASCEPT offers many advantages over comparable methods. For example, using PELT to detect multiple changepoints requires specifying an optimization penalty while ASCEPT allows an investigator to specify a significance level, a more intuitive statistical parameter. Additionally, ASCEPT is specifically designed for mHealth data, which is not true of comparable methods like CBS. For instance, CBS uses a permutation test to obtain p-values for changepoints,[7] but this approach has difficulty capturing segments containing only one observation, a feature we observed in the mHealth data from the Precision VISSTA study (**Figure 1A**). ASCEPT's Monte Carlo procedure does not run into this same problem and captures single-point segments, as shown in **Figure 4B**. In addition, CBS trims changepoints using a sum of squared within-segment deviations measure,[7] while ASCEPT uses regression to directly model the linear and seasonal trends observed in mHealth data, thereby helping to differentiate between expected behaviorally driven patterns and other patterns that may be a result of technological changes.

Importantly, ASCEPT allows an investigator to identify potential technological changepoints automatically, rather than manually. For example, while Fitbit lists previous firmware versions online, it does not readily provide release dates or specific notes regarding each one.[16] Instead, a researcher needs to manually read through online community forums for details.[17] In our investigation of these online notes, we found that update rollouts and glitches often occur over days or weeks, making it difficult to precisely determine when the data will reflect these changes. Furthermore, some changes may not even be publicized, rendering a manual search useless. In contrast, ASCEPT provides an effective way to precisely identify when technologically driven changes occur.

We note, however, that ASCEPT also has some potential limitations. First, since it involves a Monte Carlo method, ASCEPT does not guarantee the same results over repeated runs; however, using a large number of simulations mitigates this issue. ASCEPT is also computationally intensive, but its implementation is parallelized for improved performance. In addition, ASCEPT assumes that the observations are normally distributed, which may not always be true. However, normality is appropriate to use in many scenarios, such as when using the sample mean or median of a variable,[18] as we did in our application of ASCEPT to mHealth data from the Precision VISSTA study. Lastly, ASCEPT requires the selection of two thresholds: a significance level and a trimming threshold. While these parameters are intuitive, we recommend that investigators consider different trimming thresholds to select a value that is appropriate for their data. In our analysis, we found that the changepoints identified by ASCEPT were robust across a wide range of trimming threshold values.

ASCEPT also has many strengths. For instance, while we developed ASCEPT for mHealth data and tested it on data from the Precision VISSTA study, the approach is highly generalizable. For example, it could be applied to select mean-shift changepoints in any univariate time series for which linear trends and seasonality induce nuisance changepoints. Additionally, instead of only applying ASCEPT to population-level data to identify technological changepoints, an investigator could apply ASCEPT to an individual's time series to identify behavioral shifts that are not associated with broader seasonal or linear patterns. ASCEPT could also be updated to identify and remove nuisance changepoints within other trends, such as quadratic trends, which may be more prevalent in other types of data.[19] The assumptions made by ASCEPT, such as normality, could also be adjusted to allow for other distributional assumptions. In addition, while the current presentation of ASCEPT uses PELT, the processes described here could, in theory, be applied to other changepoint detection algorithms.

We designed ASCEPT as a formal process to select relevant changepoints among those proposed by PELT by modeling trends that are commonly associated with nuisance changepoints. We believe identifying these types of changepoints will be a critical step for effectively analyzing mHealth data, which often contains changepoints both from sudden changes in the propriety algorithms used to record measurements and from changes in human behavior. ASCEPT automates this process and only requires selecting two intuitive parameters. This gives it a distinct advantage over using other methods or performing a manual identification of technological changepoints, supporting its broad applicability to mHealth data analysis.

FUNDING

MQ, AC, and KG report funding from R01EB025024. MQ also reports funding from T32HL007427-39.

AUTHOR CONTRIBUTIONS

AC and KG conceptualized the project. MQ and KG developed the method. MQ wrote the software and performed the formal analysis and investigation. KG and AC curated the data and resources for the project. MQ wrote the original draft. All authors reviewed and edited the manuscript. KG supervised the project. AC and KG acquired the funding.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

ASCEPT's code and the simulated data are available online at https://github.com/matthewquinn1/changepointSelect. Those interested in the availability of the Precision VISSTA data should contact AC.

1 Kumar S, Nilsen WJ, Abernethy A, *et al.* Mobile Health Technology Evaluation: The mHealth Evidence Workshop. *Am J Prev Med* 2013;**45**:228–36. doi:https://doi.org/10.1016/j.amepre.2013.03.017

2 Silva BMC, Rodrigues JJPC, Díez I de la T, *et al.* Mobile-health: A review of current state in 2015. *J Biomed Inform* 2015;**56**:265–72. doi:https://doi.org/10.1016/j.jbi.2015.06.003

3 Kelli HM, Witbrodt B, Shah A. THE FUTURE OF MOBILE HEALTH APPLICATIONS AND DEVICES IN CARDIOVASCULAR HEALTH. *Eur Med J Innov* 2017;:92–7.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5298843/

4 Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. *Signal Process* 2020;**167**:107299. doi:https://doi.org/10.1016/j.sigpro.2019.107299

5 Killick R, Fearnhead P, Eckley I. Optimal Detection of Changepoints With a Linear Computational Cost. *J Am Stat Assoc* 2012;**107**:1590–8. doi:10.1080/01621459.2012.737745

6 Haynes K, Eckley IA, Fearnhead P. Computationally Efficient Changepoint Detection for a Range of Penalties. *J Comput Graph Stat* 2017;**26**:134–43. doi:10.1080/10618600.2015.1116445

7 Olshen A, Venkatraman ES, Lucito R, *et al.* Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004;**5**:557–72. doi:10.1093/biostatistics/kxh008

8 Chung A, Gotz D, Kappelman M, *et al. Precision VISSTA: Enabling Precision Medicine through the Development of Quantitative and Visualization Methods*. http://precisionvissta.web.unc.edu/

9 Killick R, Eckley IA. changepoint: An R Package for Changepoint Analysis. *J Stat Softw* 2014;**58**:1–19.http://www.jstatsoft.org/v58/i03/

10 Lynch G, Guo W, Sarkar SK, *et al.* The control of the false discovery rate in fixed sequence multiple testing. *Electron J Stat* 2017;**11**:4649–73. doi:10.1214/17-EJS1359

11 Seshan VE, Olshen A. *DNAcopy: DNA copy number data analysis*. 2019.

12 *Alta HR Firmware Release - 26.62.6*. 2017. https://community.fitbit.com/t5/Alta-HR/Alta-HR-Firmware-Release-26-62-6/td-p/2119538

13 Kosecki D. *New Fitbit Features Deliver Data Previously Only Available Through a Sleep Lab*. 2017. https://blog.fitbit.com/sleep-stages-and-sleep-insights-announcement/

14 *Charge 2 Sleep Stages*. 2017. https://community.fitbit.com/t5/Charge-2/Charge-2-Sleep-Stages/td-p/1907433

15 *Received Classic Sleep rather than Sleep Stages*. 2017. https://community.fitbit.com/t5/Blaze/RESOLVED-9-3-Received-Classic-Sleep-rather-than-Sleep-Stages/td-p/2174227

16 *What's changed in the latest Fitbit device update?* https://help.fitbit.com/articles/en_US/Help_article/1372

17 *Community*. https://community.fitbit.com/

18 Rider PR. Variance of the Median of Small Samples from Several Special Populations. *J Am Stat Assoc* 1960;**55**:148–50.http://www.jstor.org/stable/2282186

19 David I. Harvey, Stephen J. Leybourne, A. M. Robert Taylor. Testing for Unit Roots and the Impact of Quadratic Trends, with an Application to Relative Primary Commodity Prices. *Econom Rev* 2011;**30**. doi:10.1080/07474938.2011.553561

Figure Captions

Figure 1. The ASCEPT workflow. (A) The daily median deep sleep from the Precision VISSTA study. (B) ASCEPT broken down by stage and applied to simulated data. The first row shows the original simulated time series. The second row shows significant changepoints being iteratively identified. The third row shows changepoints within linear and seasonal trends being iteratively trimmed. The fourth row shows the simulated time series with the final set of identified changepoints. (C) The same results as B but for the deep sleep data.

Figure 2. The process for assessing the significance of new changepoints in ASCEPT. (A) The simulated data with initial changepoints $\mathcal{T}_k = \{305, 600\}$. The log-likelihood, assuming independent and identically distributed observations within-segment, is -3727.3. (B) The simulated data set with the next set of changepoints $\mathcal{T}_{k+1}^* = \{49, 60, 305, 600\}$. The log-likelihood is -3512.3, thus the observed change in the log-likelihood is 215.0. (C) A Monte Carlo sample with the initial changepoints at {305, 600} shown. The observations in each segment are randomly drawn from a normal with a mean and standard deviation equal to that for the corresponding segment in subfigure A. The log-likelihood is -3746.6. (D) The same Monte Carlo sample from subfigure C, but now with the next set of changepoints at {49, 60, 305, 600} shown. The log-likelihood is -3745.0. The change in the log-likelihood for this Monte Carlo sample under the null is therefore 1.6. The process in subfigures C and D are repeated a large number of times to generate an empirical null distribution for the change in the log-likelihood. In all plots, the segments between the identified changepoints are numbered.

Figure 3. The trimming process in ASCEPT. (A) The simulated data set with an initial set of true changepoints. For illustrative purposes, only a subset of the changepoints found after running the first stage of ASCEPT is shown. (B) Assessing the changepoint between segments 2 and 3, a relevant changepoint. The cross-segment fits are more than 3 times worse than the best piecewise fit. (C) Assessing the changepoint between segments 4 and 5, a nuisance changepoint due to a linear trend. The cross-segment linear fit is only about 3% worse than the best piecewise fit. (D) Assessing the changepoint between segments 8 and 9, a nuisance changepoint due to seasonality. The cross-segment harmonic fit is only about 9% worse than the best piecewise fit.

Figure 4. Overall results from applying ASCEPT to (A) the simulated data, as well as mHealth data from the Precisions VISSTA study measuring (B) median deep sleep, (C) median light sleep, and (D) median total sleep.

Figure 5. Comparison of ASCEPT with CBS when applied to (A) the simulated data, as well as mHealth data from the Precisions VISSTA study measuring (B) median deep sleep and (C) median times woken during the night.

Figure 6. Illustration of applying a simple correction process to simulated data after identifying changepoints using either ASCEPT or CBS. (A) The best model fits using ASCEPT changepoints. (B) The best model fits using CBS changepoints. (C) The corrected series using ASCEPT changepoints. (D) The corrected series using CBS changepoints.