SensibleSleep: A Bayesian Model for Learning Sleep Patterns from Smartphone Events

Andrea Cuttone[1]�*, Per Bækgaard[1]�, Vedran Sekara[1,3], Håkan Jonsson[3], Jakob Eg Larsen[1], Sune Lehmann[1,2]

1 DTU Compute, Technical University of Denmark, Kgs. Lyngby, Denmark, **2** The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark, **3** Sony Mobile, Nya Vattentornet, Mobilvägen, Lund, Sweden

�he These authors contributed equally to this work.
* ancu@dtu.dk

Abstract

We propose a Bayesian model for extracting sleep patterns from smartphone events. Our method is able to identify individuals' daily sleep periods and their evolution over time, and provides an estimation of the probability of sleep and wake transitions. The model is fitted to more than 400 participants from two different datasets, and we verify the results against ground truth from dedicated armband sleep trackers. We show that the model is able to produce reliable sleep estimates with an accuracy of 0.89, both at the individual and at the collective level. Moreover the Bayesian model is able to quantify uncertainty and encode prior knowledge about sleep patterns. Compared with existing smartphone-based systems, our method requires only screen on/off events, and is therefore much less intrusive in terms of privacy and more battery-efficient.

Introduction

Sleep is an important part of life, and quality of sleep has a significant impact on individual well-being and performance. This calls for methods to analyze sleep patterns in large populations, preferably without laborious or invasive consequences, as people typically disapprove of the use of intrusive technologies [1].

Large scale studies of human sleep patterns are typically carried out using questionnaires, a method that is known to be unreliable. It is possible to perform more accurate studies, but these are currently carried out within small controlled environments, such as sleep labs. In order to perform accurate measurements of sleep in large populations—consisting of thousands of individuals—without dramatically increasing costs, alternative methods are needed.

Smartphones have become excellent proxies for studies of human behavior [2, 3], as they are able to automatically log data from built-in sensors (GPS, Bluetooth, WiFi) and on usage patterns (phone calls, SMS and screen interaction), from which underlying user behavioral patterns can be derived.

Smartphone data has been used to infer facets of human behavior such as social interactions [4], communication [5], mobility [6], depression [7] and also sleep patterns [8]. Either paired

with additional sensors or on their own, mobile app solutions are able—sometimes very ingeniously—to track individual sleep patterns and visualize them. We cite as examples *Smart Alarm Clock* [9], *Sleep Cycle* [10], *SleepBot* [11], and *Sleep as Android* [12].

In this paper we suggest extending previous approaches, using a Bayesian model to infer rest and wake periods based on smartphone screen activity information. The advantages of our proposed Bayesian approach *SensibleSleep*, as compared to previous work, are that it:

- is less sensitive to "noisy" data, for instance infrequent phone usage during sleep interruptions (such as checking the phone at night)

- is able to quantify not only specific rest and wake times but also characterize their distributions and thus uncertainty

- can encode specific prior beliefs, for instance on expected rest periods (when desirable)

- can capture complex dependencies between model variables, and possibly even detect and relate patterns that are common to a group of people with diverging individual patterns (when using one of the proposed hierarchical models), such as detecting how available daylight may modulate sleep patterns across an otherwise heterogeneous group of users

Our method, moreover, only needs screen on/off events and is thus *non-intrusive*, *privacy-preserving*, and has *lower battery cost* than microphone or accelerometer based ones.

Although dedicated sleep trackers or fitness tracking bands can provide much more precise and fine-grained data on sleep patterns, their adoption is still quite small in comparison with mobile phones. Therefore the data potentially available from smartphones can enable the analysis of sleep patterns at a very large scale, much larger than using data from dedicated sleep tracking devices.

We start by providing an overview of the related work. We then describe the collected data, and introduce the Bayesian model. We compare the model results with ground truth obtained by sleep trackers, and show how the model is able to infer the sleep patterns with high accuracy. Finally we describe the individual and collective sleep patterns inferred from the data.

Related Work

A key finding by Zhang et al. [13] shows a global prevalence of sleep deprivation in a group of students, partly linked to heavy media usage. In this study sleep patterns are largely deduced from the teachers' perception or based on individual self-reports, lacking more direct measurements.

Corroborating this finding, Orzech et al. [14] report that digital media usage before bedtime is common among university students, and negatively impacts sleep. The findings are based on studies involving self-reports through (online) sleep diaries and digital media surveys, and also lacks more direct measurements of sleep patterns. Additionally, this would make it possible to increase the scale of the experiment and enable the study of larger populations.

Abdullah et al. [8] have previously demonstrated using 9 subjects how a simple rule-based algorithm is able to infer sleep onset, duration and midpoint based on a (filtered) list of screen on-off patterns with the help of previously learned individual corrective terms, and further analyzed behavioral traits of the inferred *circadian rhythm* [15, 16]. The algorithm uses an initial two weeks of data with journal self-reported sleep for learning key corrective terms in order to improve the accuracy and compensate for differences between *actual* sleep and *inferred* nightly rest period. The method has been verified against a daily online sleep journal and results in differences less than 45 minutes of average sleep duration over the entire analysed period. While our proposed Bayesian model, which has been applied to more than 400 users, may be more

complex, it increases the robustness and allows us to better quantify the uncertainties of the inferred resting periods as well as offer the possibility of building more advanced models across heterogeneous groups of users. In particular, our model may better be able to handle short midnight interruptions, which appear to be not uncommon, without any additional filtering.

In contrast to Abdullah et al. using (only) screen on-off events, a fine-grained sleep monitoring by "hearing" and analyzing breathing through the earphone of a smartphone is suggested by Ren et al. [17]. Here six users tested the system over a period of 6 months, demonstrating the feasibility of using smartphones for the purpose of analysing breathing patterns, using a Respiration Monitor Logger as ground truth. Sleep estimates are not directly inferred in this paper, however. This technology is also non-invasive, although it does requires capturing and analyzing large samples of audio data.

iSleep [18] proposes detecting sleep patterns by means of a decision tree model, also based on audio features. The system was evaluated with 7 users for a total of 51 days, and shows high accuracy in detecting snoring and coughing as well as sleep periods, but report drops in performance due to ambient noise.

Increasing the number of features, the *Best Effort Sleep model* [19] is based on a linear combination of phone usage, accelerometer, audio, light, and time features using a self-reporting sleep journal, and subsequently achieved a 42 minutes mean error on 8 subjects in a test period of 7 days.

Other work also tries to estimate sleep quality, for example *Intelligent Sleep Stage Mining Service with Smartphones* [20], which uses Conditional Random Fields on a similar set of features trained on 45 subjects over 2 nights, and reports over 65% accuracy of detection of sleep phases, compared to EEG ground truth on 15 test subjects over 2 nights.

Candy Crushing Your Sleep [21] uses the longest period of phone usage inactivity as heuristic for sleep, with some ad-hoc rules for merging multiple periods, and proceeds to quantify the sleep quality and to identify aspects of daily life that may affect sleep. The inferred sleep period was however not validated against any ground truth.

The *Sleep Well* framework [22] deploys a Bayesian probabilistic change-point detection, in parallel with an unsupervised classification, of features extracted from accelerometer data, in order to identify fine-grained sleep state transitions. It then uses an active learning process to allow users to incrementally label sleep states, improving accuracy over time. It was evaluated both on existing datasets with clinical ground truth, and on 17 users for 8-10 days with user diary data as ground truth, reaching an average sleep stage classification accuracy approaching 79%.

In comparison, even though sleep quality is not estimated, our non-intrusive model only needs screen on/off events and has been tested on a large user-base, and can suitable for very large-scale deployment.

Methods

Data Collection

We have analyzed two datasets in this work.

The first dataset (A) was provided by Sony Mobile, and contains smartphone app launches coupled with sleep tracking data from the SWR10 and SWR30 fitness tracking armbands [23]. For each user we have a set of records containing an anonymized unique user identifier, a timestamp and the unique app package name. Note that the model only uses the app launch timestamp and completely ignores the app identifier, therefore no privacy risks related to app names are present. The sleep tracking data indicates when each user is detected asleep or awake with a granularity of one minute, serving as ground truth that we will compare our

results against. From this dataset we select 126 users that have at least 3 hours of tracked sleep per day, and have between 2 and 4 weeks of contiguously tracked sleep.

The second dataset (B) originates from the *SensibleDTU* project [24], which collected smartphone sensor data for more than 800 students at the Technical University of Denmark. In this dataset we focus on the screen interaction sensor that records whenever the smartphone screen is turned on or off, either by user interaction or by notifications. Each record contains a unique user identifier, a timestamp, and the event type (on or off). From this dataset we select 324 users in November 2013 that have at least 10 events per day, thus filtering out users with gaps in the collected data or with very sparse data. There is on average ≈ 76 screen-on activations pr. day pr. user in this period.

Data collection for the SensibleDTU dataset was approved by the Danish Data Protection Agency, and written informed consent has been obtained for all study participants. Data collection for the Sony dataset has been approved by the Sony Mobile Logging Board and written informed consent has been obtained for all study participants according to the Sony Mobile Application Terms of Service and the Sony Mobile Privacy Policy.

Model Assumptions

The underlying assumptions of the model are (1) that the user is in one of two modes: being *awake* or *sleeping*, and (2) that mobile phone usage differs between the two modes. In particular a user will have many screen interactions when awake, and very few or even no interactions when sleeping.

Sleeping is here considered as an extended *resting* period that typically takes place once every 24 hours at roughly similar times, as governed by the users circadian rhythm and influenced by socio-dynamic structures, during which the owner physically rests and/or sleeps. Resting periods, however, might be interrupted by short periods of activity, such as checking the time on the phone or responding to urgent messages. This behavior leads to two different activity levels, which we label λ_{awake} and λ_{sleep}, one for each mode.

If we can deduce when the switchpoint between the two distributions occur during each 24 hour period, we can also infer the time during which the owner is *resting* for the night, and thereby also the period within which sleeping takes place.

Short of using the more invasive EEG or polysomnographic methods, properly differentiating the resting period and actual sleep is difficult; even sleep diaries may easily contain reporting bias or be somewhat inaccurate. To remove self-reporting bias and to study a larger population we have therefore decided on using a motion-based detector (Sony fitness tracking armbands) as ground truth.

If higher accuracy would be required, applying individual corrective terms (i.e. average sleep/rest time differences) learned from an initial period by more accurate means (polysomnography, external observer or possibly a careful user diary) might be possible, similar to what as demonstrated by Abdullah et al. [8].

Model Structure

Each user is considered independently. We divide time into 24–hour periods starting at 16:00 and ending at 15:59 on the next calendar day, so that the night period and the expected sleep midpoint is in the middle, for convenience. Each day is divided into $n = 24 * 4 = 96$ time bins of size 15 minutes. We count the number of events that start within each time bin, where an event is an app launch for dataset A and a screen-on for dataset B. Information about the duration of the events is purposely discarded, as phone usage typically takes place in short bursts. This is supported by the median duration of screen events in dataset B, which is ≈ 26.5 seconds.

A reasonable weakly informative prior assumption is that the count of events k in each time bin follows a Poisson distribution (additional comments on this assumption are available in S1 Appendix):

$$P(k) = \text{Poisson}(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

with $\lambda = \lambda_{awake}$ or $\lambda = \lambda_{sleep}$, depending on the mode of the user. It is, furthermore, assumed that the user mode, and consequently the value for λ, is determined by two switchpoint variables t_{sleep} and t_{awake}, both assuming values from 0 to n:

$$\lambda = \begin{cases} \lambda_{sleep} & \text{if } t_{sleep} \leq t < t_{awake} \\ \lambda_{awake} & \text{if } t < t_{sleep} \vee t \geq t_{awake} \end{cases}$$

For simplicity, all models assume that λ_{sleep} is identical for all days of a given user. It can be expected that users have a very low number of screen events during sleep mode, which is encoded in this prior belief:

$$\lambda_{sleep} \sim \text{Exponential}(10^4)$$

Here Exponential represents the exponential distribution:

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The rate parameter is set to a very large value to encode our prior belief that almost no events should happen during the sleep time.

Fig 1 shows an illustration of the model idea.

We now propose five different models, which differ in the assumptions made on the relation of the rate and sleep/awake time parameters for different days.

Pooled-Pooled Model: Pooled Times and Rates

The simplest model assumes that for a given user there is a single λ_{awake}; i.e. the user has very similar phone interaction patterns each day. Also t_{sleep} and t_{awake} are each identical for all days, that is: the user goes to sleep, and wakes up, at the same times each day:

$$\begin{aligned} t_{sleep} &\sim \text{DiscreteUniform}(0, n) \\ t_{wake} &\sim \text{DiscreteUniform}(0, n) \\ \lambda_{awake} &\sim \text{Gamma}(2.5, 1) \end{aligned}$$

Here DiscreteUniform$(0, n)$ represents a uniform probability to choose a timebin between 0 and $n = 96$. No additional prior knowledge of t_{sleep} and t_{awake} is assumed; there is equal probability of any bin value. In other words, sleep and awake time are equally probable at any time of the day. The prior for λ_{awake} is chosen to represent our prior belief of a reasonable rate of events, specifically with both mean and variance = 2.5 (events/bin) and a longer tail than a normal distribution.

Independent-Pooled Model: Independent Times

A somewhat more realistic model would assume that each day has independent t_{sleep} and t_{awake} times, while still sharing λ_{awake} rates. Therefore in this model there are t^i_{sleep} and t^i_{awake}, with
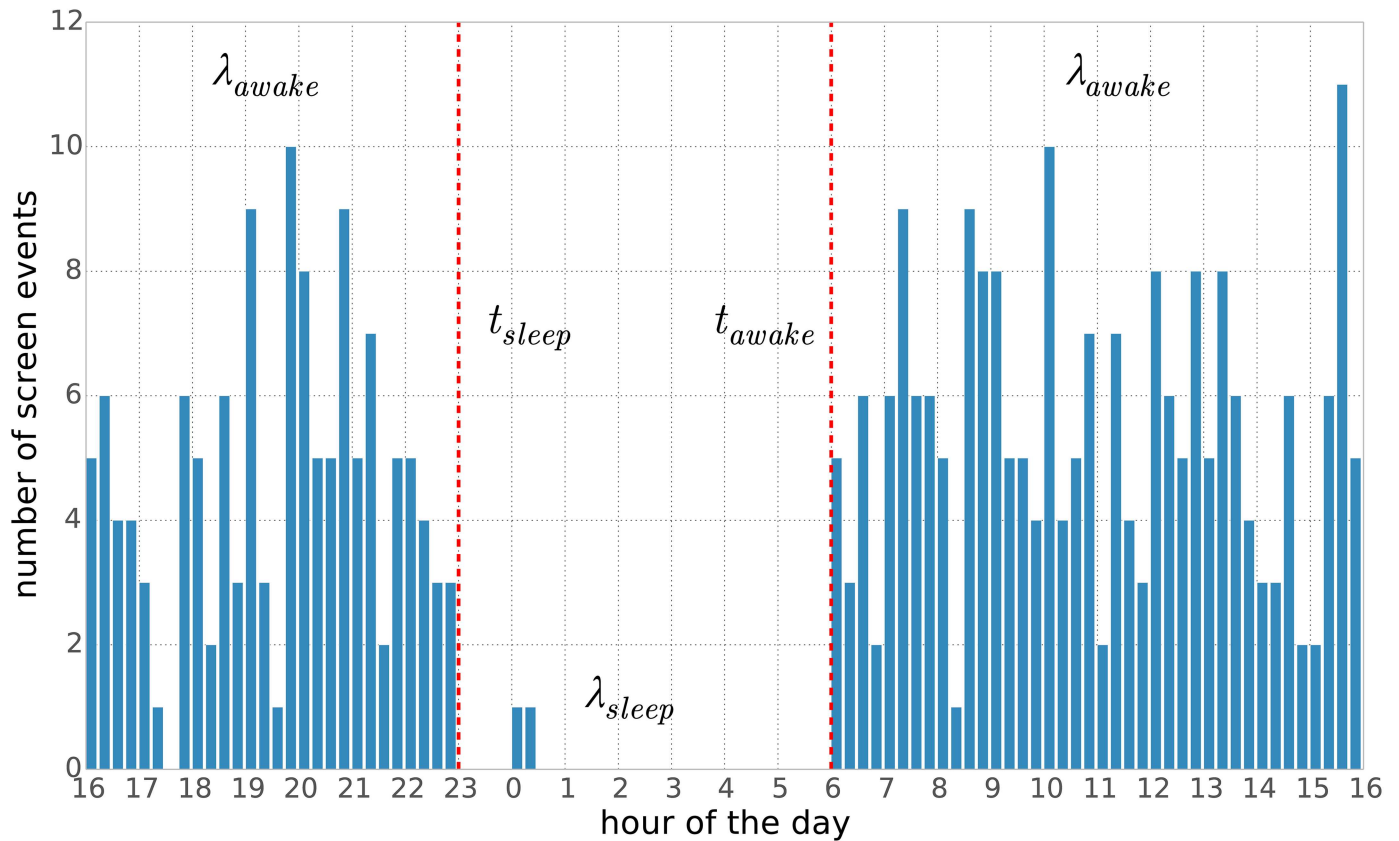
Fig 1. Conceptual illustration of the model. We assume that for each day the event counts follow two different Poisson distributions: one for sleep periods (rate λ_{sleep}) and one for awake periods (rate λ_{awake}). Furthermore we assume that two switchpoints t_{sleep} and t_{awake} determine the rate (i.e. the Poisson distribution) that generates the events.

$i = 1 \ldots m$, one for the each of the considered days:

$$t_{sleep}^{i} \sim \mathrm{DiscreteUniform}(0, n) \text{ for } i = 1 \ldots m$$
$$t_{wake}^{i} \sim \mathrm{DiscreteUniform}(0, n) \text{ for } i = 1 \ldots m$$
$$\lambda_{awake} \sim \mathrm{Gamma}(2.5, 1)$$

The rest of the model remains as above.

Independent-Independent Model: Independent Times and Rates

It may further be assumed that each day could have its own specific activity rate. We modeled this as separate λ_{awake}^{i} for each of the m days, in addition to t_{sleep} and t_{awake} for each of the m days:

$$t_{sleep}^{i} \sim \mathrm{DiscreteUniform}(0, n) \text{ for } i = 1 \ldots m$$
$$t_{wake}^{i} \sim \mathrm{DiscreteUniform}(0, n) \text{ for } i = 1 \ldots m$$
$$\lambda_{awake}^{i} \sim \mathrm{Gamma}(2.5, 1) \text{ for } i = 1 \ldots m$$

Independent-Hyper Model: Hierarchical Rates

The assumption that each day's interaction rate is completely independent may not be correct. It may not be unreasonable to imagine that the daily rate(s) arise from an underlying user-specific rate; i.e. the user may have certain habits that varies from day to day but share some similarities specific to that user. This is modeled by adding α_λ and β_λ hyperparameters to the Gamma priors for λ_{awake}^i:

$$
\begin{aligned}
t_{sleep}^i &\sim \text{DiscreteUniform}(0, n) \text{ for } i = 1 \ldots m \\
t_{wake}^i &\sim \text{DiscreteUniform}(0, n) \text{ for } i = 1 \ldots m \\
\alpha_\lambda &\sim \text{Exponential}(1) \\
\beta_\lambda &\sim \text{Exponential}(1) \\
\lambda_{awake}^i &\sim \text{Gamma}(\alpha_\lambda, \beta_\lambda) \text{ for } i = 1 \ldots m
\end{aligned}
$$

We do not have strong prior beliefs for α and β, so we set their prior distributions to generic exponential distribution with rate parameter = 1, Exponential(1).

Hyper-Hyper Model: Hierarchical Times and Rates

Finally we could assume that each day's sleep and awake times derive from an underlying circadian rhythm that is specific to the user, but still modulated by events that take place during the week. This can be modeled by changing the t_{sleep}^i and t_{awake}^i priors to a normal distribution, with hyperparameters α_t, β_t and τ_t as follows:

$$
\begin{aligned}
\alpha_t &\sim \text{Exponential}(1) \\
\beta_t &\sim \text{Exponential}(1) \\
\tau_t &\sim \text{Gamma}(\alpha_t, \beta_t) \\
t_{sleep}^i &\sim \text{Normal}(8 * (n/24), \tau_t) \text{ for } i = 1 \ldots m \\
t_{wake}^i &\sim \text{Normal}(15 * (n/24), \tau_t) \text{ for } i = 1 \ldots m \\
\alpha_\lambda &\sim \text{Exponential}(1) \\
\beta_\lambda &\sim \text{Exponential}(1) \\
\lambda_{awake}^i &\sim \text{Gamma}(\alpha_\lambda, \beta_\lambda) \text{ for } i = 1 \ldots m
\end{aligned}
$$

The t_{sleep}^i are here chosen to be centered at the bin corresponding to 23:00, while the t_{awake}^i are centered at the bin corresponding to 07:00. Also in this case we have no strong prior knowledge of the τ_t, α_t and β_t parameters, so we set their prior distribution to a non-informative Exponential and Gamma respectively.

Model Fitting and Selection

The models are fitted using Markov Chain Monte Carlo (MCMC) sampling [25], where the parameter values are estimated by a random walk in the parameter space guided by the log likelihood. We use the *pymc3* python library [26, 27] for running the sampling, but any MCMC framework could be used to implement our model. The result of the Bayesian inference is a trace that captures the most probable values of the parameters, and also gives an indication of the uncertainty of the estimation.

It is important to note that the models are unsupervised, which means that they are fitted only to the number of events without having access to the ground truth of the actual sleep

patterns. This allows the model to be fit to other datasets where we do not have ground truth of sleep patterns, which is desirable if the sleep inference has to be deployed on a large scale. For dataset A we verify the fit by comparing with the sleep patterns from sleep trackers, while for dataset B we evaluate the fit by inspecting the inferred sleep patterns.

In order to find the model that provides the best overall fit for the intended purpose without introducing too many degrees of freedom, we compare the log posterior from the traces of the models, logp, and see how they converge.

One example of a plot of logp traces for the five models is shown in Fig 2, which shows that the hyper-hyper model (blue) has the highest (least negative) logp, followed by the independent-hyper model for dataset B. The three other models appear with lower logp. In 76% of the analyzed cases of dataset A (84% for dataset B), the hyper-hyper model has the highest logp score, followed by the independent-hyper model with the highest logp in 11% (13%) of the cases.

The logp estimation does not, however, take into account the added complexity of the more advanced models. An attempt to do so is the Deviance Information Criterion (DIC) [28], which penalizes the increased degrees of freedom (more model parameters) that usually result in a model that is easier to fit to the data. Fig 3 shows the Relative DIC score (vs. the simplest model, pooled-pooled). The order is identical for both datasets.

Further, Table 1 compares the 5 models by ranking the calculated DIC for all 126 and 324 users. The *median* rank shows that the hyper-hyper model is the "best" model; it has a probability of being the best ranked model (*p(Best)*) in 62% of the cases for dataset A (69% for dataset B). The independent-hyper model follows as a somewhat distant 2nd best, ranking highest in 17% (19%) of the cases. For further validation of the goodness of fit of the model, see S2 Appendix.

It should be noted that, in addition to their different abilities to reflect the underlying assumptions and provide varying levels of fit to the actual data, the models also differ in their runtime; the most complex model typically takes 15 times longer to execute than the simplest. In particular, the hyper-hyper model on average had a runtime that is 60% longer than the independent-hyper model, so there may be cases where the latter would be a better model to use despite the slightly worse DIC ranking.

Results

All five models have been run on both datasets, producing an estimation of the times of sleep and wake up for each day, as well as estimates for the other hyperparameters, for each user. Moreover, we calculated logp and DIC as discussed in the previous section. We firstly verify the accuracy our method using the ground truth from the sleep trackers. We then provide a qualitative analysis of some key examples of individual sleep patterns, and a description of the aggregated sleep patterns for both datasets. For the remainder of the paper we restrict our analysis to the model with the best fit, the hyper-hyper model.

Comparison to Related Work and to Ground Truth

To assess the results, we compare the sleep periods inferred by our model and those inferred by a previously suggested rule-based method to the ground truth collected by the Sony sleep trackers.

For each day we calculate the time of sleep and time of awake inferred by our model as the mean of the t^i_{sleep} and t^i_{wake} respectively, and we consider the user asleep (Z = 1) for all time bins between t^i_{sleep} and t^i_{wake}, and awake (Z = 0) for the remaining bins.
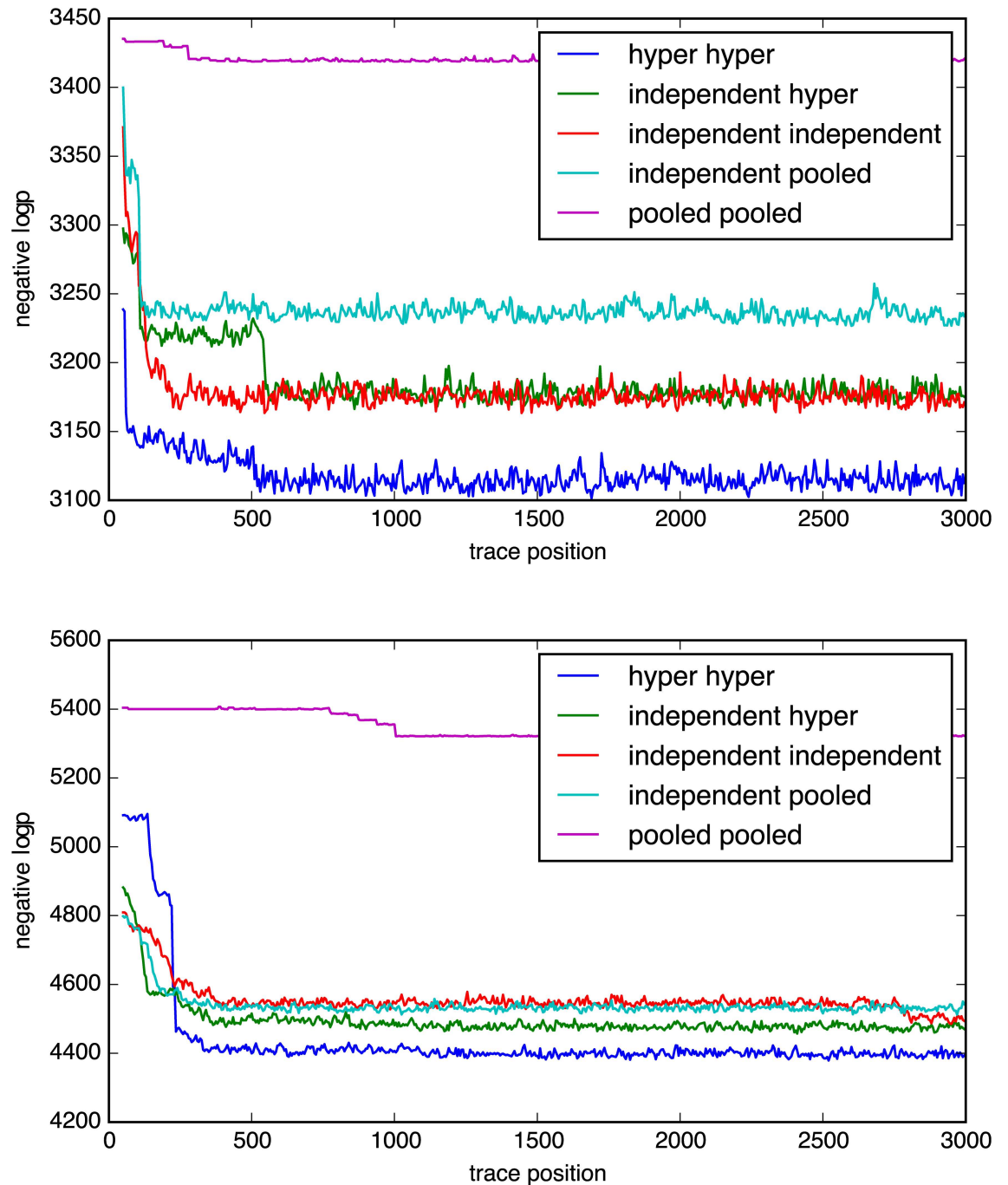
Fig 2. Typical logp traces (A top, B bottom).

For a representative and comparable method, we chose to implement a rule-based algorithm similar to what is proposed by Abdullah et. al. [8] to derive sleep data for dataset A. This rule-based method essentially works by finding the longest contiguous sleep period, with a prior assumption that sleep must start after 10 PM and before 7 AM next morning. Note that the original algorithm is based on screen on-off events and furthermore discards events of
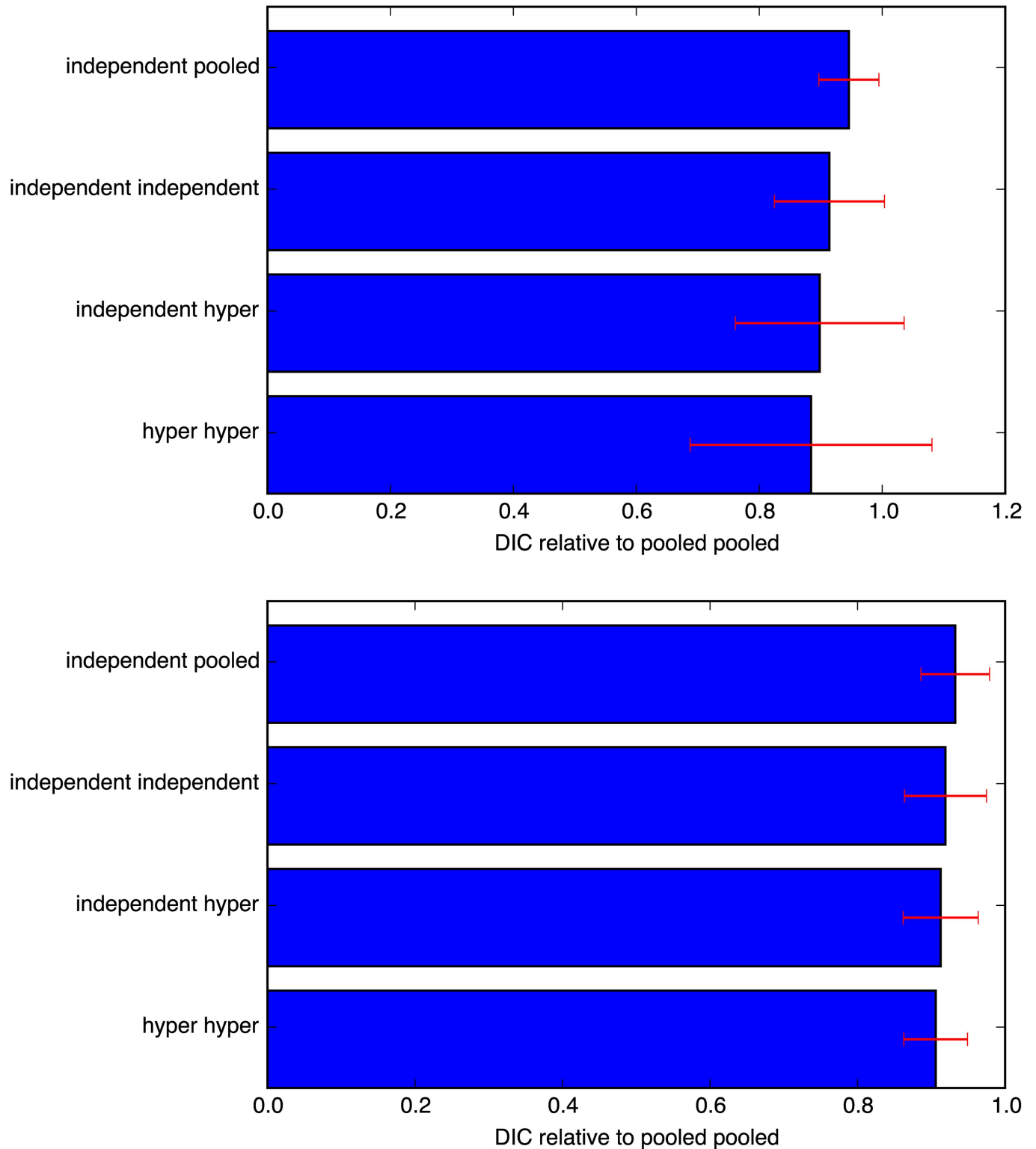
Fig 3. Relative DIC scores (A top, B bottom), sorted by their mean value (error bars represent one standard deviation). For both datasets the order is the same, with the hyper-hyper model having the lowest mean DIC.

doi:10.1371/journal.pone.0169901.g003

Table 1. Model DIC comparisons.

	Model Ranks	Median	Mean		p(Best)	Mean Relative DIC	
			Value	(StdDev)		Value	(StdDev)
A	pooled-pooled	5	4.27	(1.37)	0.10	0.96	(0.16)
	independent-pooled	4	3.82	(0.85)	0.03	0.95	(0.05)
	independent-independent	3	2.86	(1.08)	0.08	0.91	(0.09)
	independent-hyper	2	2.29	(0.83)	0.17	0.90	(0.14)
	hyper-hyper	1	1.76	(1.11)	0.62	0.88	(0.20)
B	pooled-pooled	5	4.70	(0.89)	0.02	0.99	(0.01)
	independent-pooled	4	3.75	(0.66)	0.02	0.93	(0.05)
	independent-independent	3	2.92	(1.02)	0.09	0.92	(0.06)
	independent-hyper	2	2.06	(0.69)	0.19	0.91	(0.05)
	hyper-hyper	1	1.56	(0.94)	0.69	0.91	(0.04)

short duration during the night; in our case we use app launches with no available duration, and thus cannot discard events of short duration.

For the sleep trackers we can directly mark each time bin as sleep ($Z = 1$) if the trackers have detected at least one sleep status in that bin, and awake ($Z = 0$) otherwise.

We again consider one user at a time. For each user we now have three binary matrices: two inferred sleep status values per time bin from either model, and one measured sleep status value per time bin (ground truth). We evaluate this as two binary classification problems, and calculate accuracy, precision, recall and F1 for each model and for each user according to the definitions:

$$\text{accuracy} = \frac{\text{correct predictions}}{\text{predictions}}$$

$$\text{precision} = \frac{\text{true positives}}{\text{predicted positives}}$$

$$\text{recall} = \frac{\text{true positives}}{\text{all positives}}$$

$$\text{F1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Fig 4 shows the resulting distribution of accuracy, precision, recall and F1 scores for the proposed method. The SensibleSleep method achieves a mean accuracy of 0.89, and a mean F1 score of 0.83. The below-average scores for some users are expected, since it is likely that among the large population under study there will be people having irregular sleep schedule or noisy sleep ground truth.

Fig 5 shows the corresponding *complementary cumulative distributions* of the accuracy, precision, recall and F1 scores of the proposed SensibleSleep model vs that of the rule-based model [8]. The results are generally comparable between the two models, on this particular dataset. Our model has slightly better accuracy and precision whereas the previously suggested rule-based model has a slightly better recall. The F1 scores, which weights precision and recall equally, are comparable.
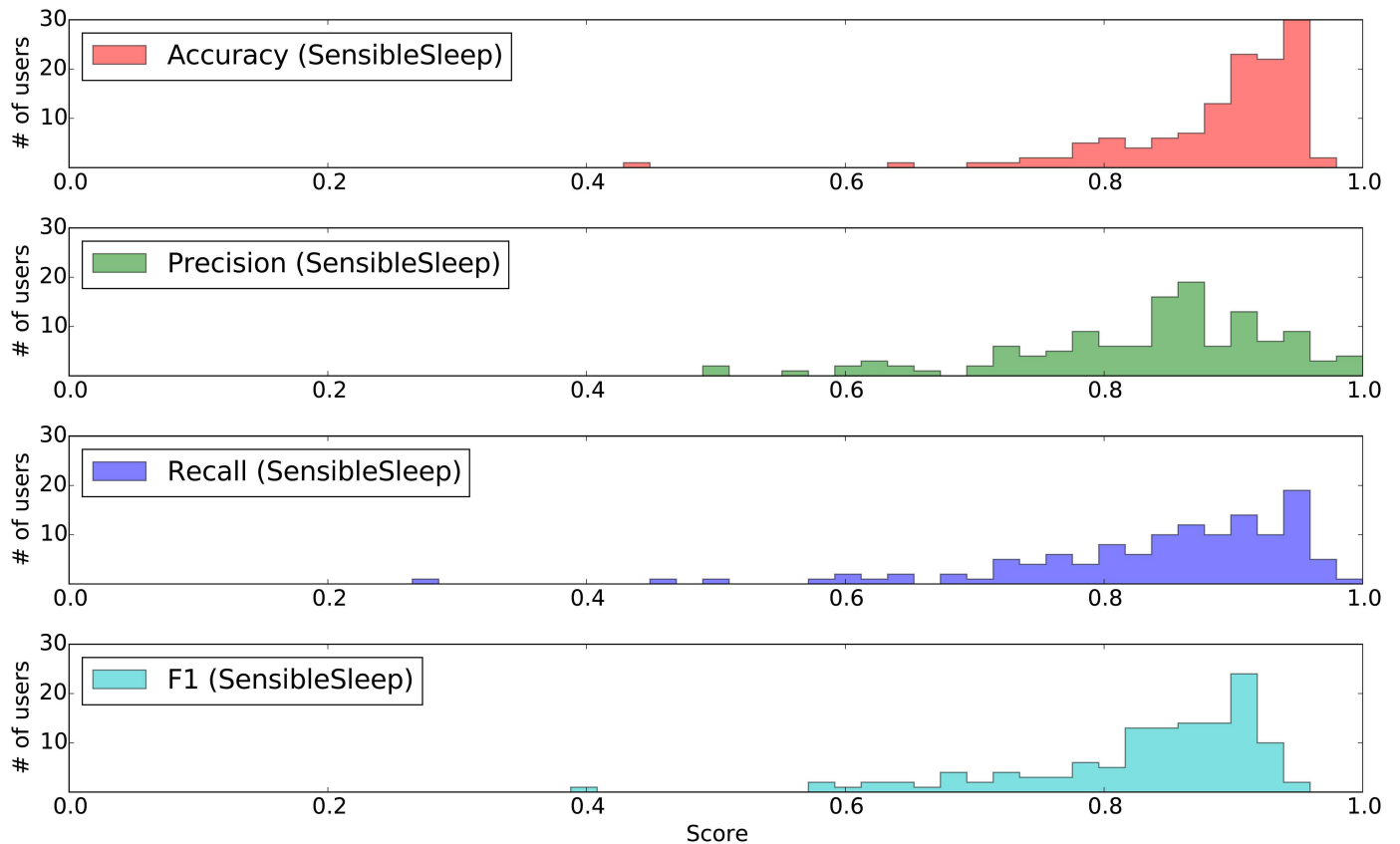
Fig 4. Histogram of the calculated accuracy, precision, recall and F1 score for users in dataset A, comparing the proposed method to the sleep tracker ground truth.

Individual Sleep Patterns

We now analyze individual sleep patterns to show the results of the model in details. For each user we create a visualization of sleep schedules. We call this the *sleep matrix*. Each row represents one day, and each column represents one time bin. The blue color shows the probability that sleep takes place within the interval; the darker the color the higher the probability. The red dots show activity count per bin; the larger the radius the more events are registered within that particular bin. This compact representation is able to capture at a glance the sleep patterns of individuals over time. We have created one such sleep matrix for each of the users, which allows us to inspect hundreds of sleep patterns quickly. Large individual variability both in sleep schedules (regular, irregular) and in phone activity (low, high, during day or night) are noticeable. Still, in most cases it is evident that the model is able to capture a reasonable sleep period, even if it may have been somewhat interrupted.

Let us consider the inferred sleep patterns for two example users in Fig 6. The top user has a pretty regular schedule, waking up around 5:30 except every few days, when he/she wakes up later—presumably due to vacation or weekends. Notice the light blue sections that indicate how the model is less confident about the probability of sleep due to events that do not follow the usual patterns. The bottom user instead has a much more unstable app usage, therefore the model infers a correspondingly more unstable sleep schedule. The bottom user has also some events in the middle of the night throughout many days (which is presumably checking the phone at night) yet the model is still able to correctly infer this being a sleep phase. Finally
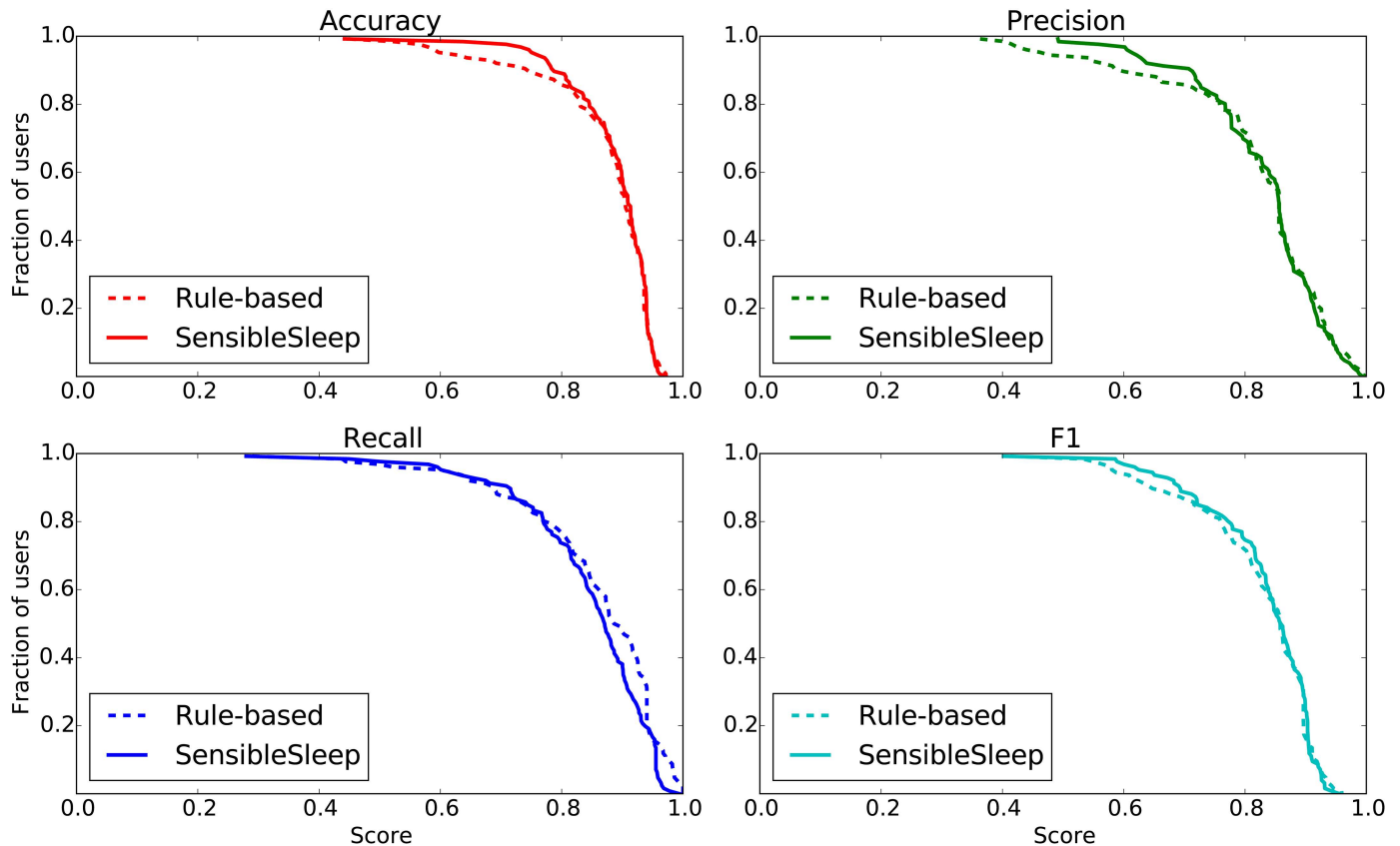
Fig 5. Complementary cumulative distribution of accuracy, precision, recall and F1 scores for users in dataset A, comparing the proposed model (solid line) to the rule-based model (dashed line), showing the proportion of users (y-axis) having a score less than or equal to a specific value (x-axis).

notice how the two users have significantly different intensity of app usage (the bottom one uses the phone much more than the top one), yet this is not a problem since the model learns individual activity rates.

Aggregated Sleep Schedules

In this section we also quantify the aggregated sleep patterns. From the posterior probability distribution functions (PDFs), $P_{t_{sleep}}(t)$ and $P_{t_{awake}}(t)$, the probability that the user is sleeping can be estimated as follows:

$$P_{sleep}(t) = P_{t_{sleep}}(t) - P_{t_{awake}}(t)$$

This is equivalent to stating that a user is currently sleeping if he has passed the time of falling asleep but has not yet passed the time waking up.

The derived values of sleep-length $t_{sleeplength}$ and mid-sleep time $t_{midsleep}$ can be calculated directly from the values of t_{sleep} and t_{awake} for each sample of the trace, and the posterior density can be estimated for these derived values in a similar way as for the model parameters. Fig 7 shows the aggregate posterior probability density functions for t_{sleep} and t_{awake} for the 126 users of dataset A over 15−30 days, and for the 324 users of dataset B over a selected period of 30 days (just after semester start).
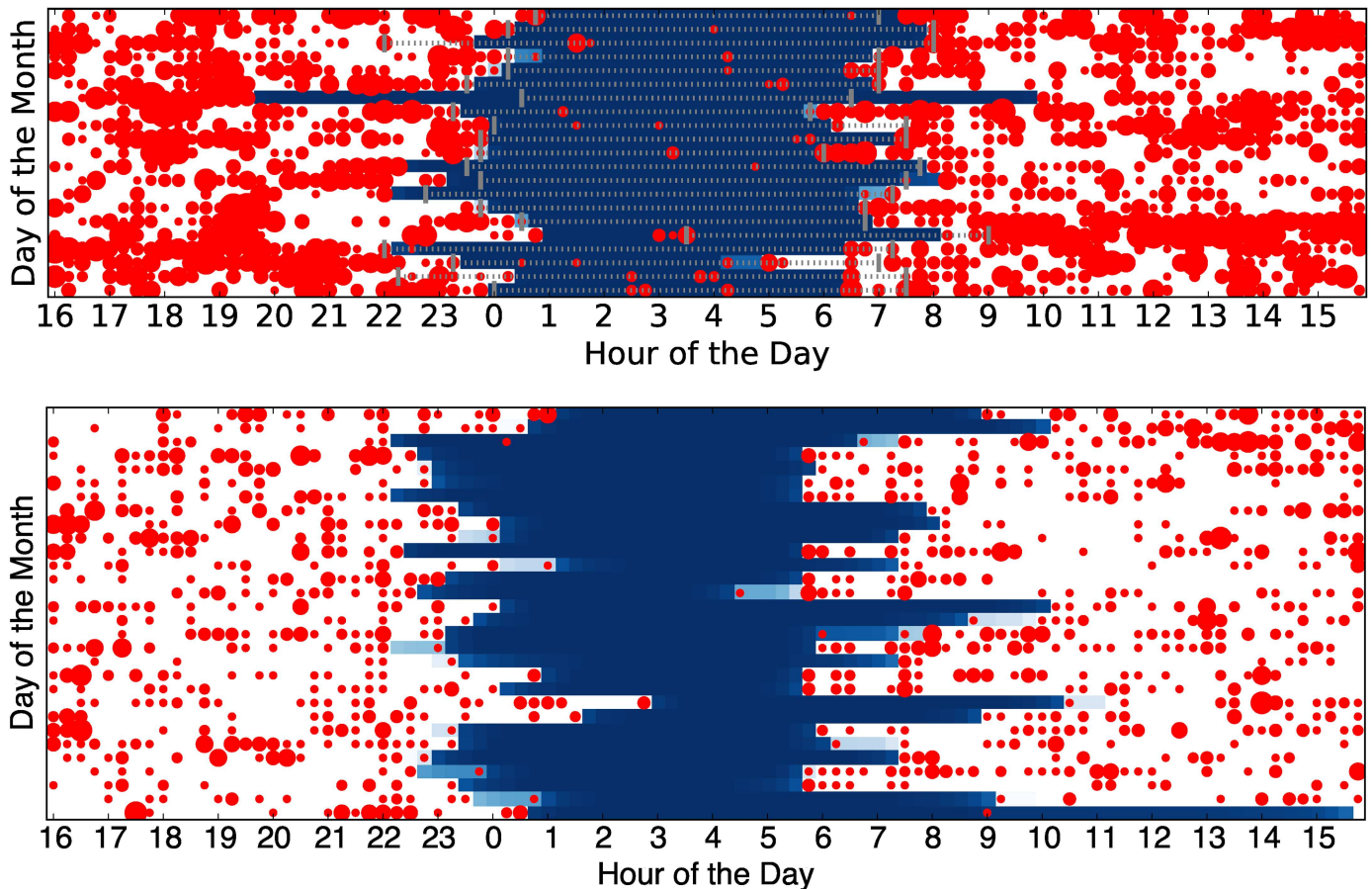
Fig 6. Sleep matrix of two sample users (21 days from dataset A top, 30 days from dataset B bottom). The ground truth derived from the sleep tracker is shown as a dotted line overlaid the matrix for dataset A.

It may not be entirely meaningful to average the sleep patterns from all users, but it serves to illustrate the distribution of t_{sleep} and t_{awake} for a larger population. Table 2 summarizes the sleep and wake times.

Across the 30 (14-28) analyzed days for the 324 (126) users of the study, the distribution of sleep durations are as shown in Fig 8. The model allows us to easily compute such metrics. The mean value is around 8:02 (±2h 36m) for dataset A and 7:20 (±2h 28m) for dataset B. Notice how the distributions are not completely similar; this is likely due to the fact that the larger dataset B captures the sleeping behavior of students as opposed to dataset A that may have a more diverse demographic distribution.

Fig 9 shows the probability density functions for the t_{sleep} and t_{awake} times for all users of dataset B, grouped according to weekday. Mondays to Thursdays appear quite similar, but Friday shows a much wider distribution; users typically go to bed much later on Friday and sleep in on Saturday. The distributions start to narrow down Saturday and Sunday but are more "week-like" only from Tuesday morning again.

Discussion

The main contribution of this work is to show how simple counts of smartphone interactions can be used to infer sleep patterns with reasonably high accuracy. We have demonstrated how
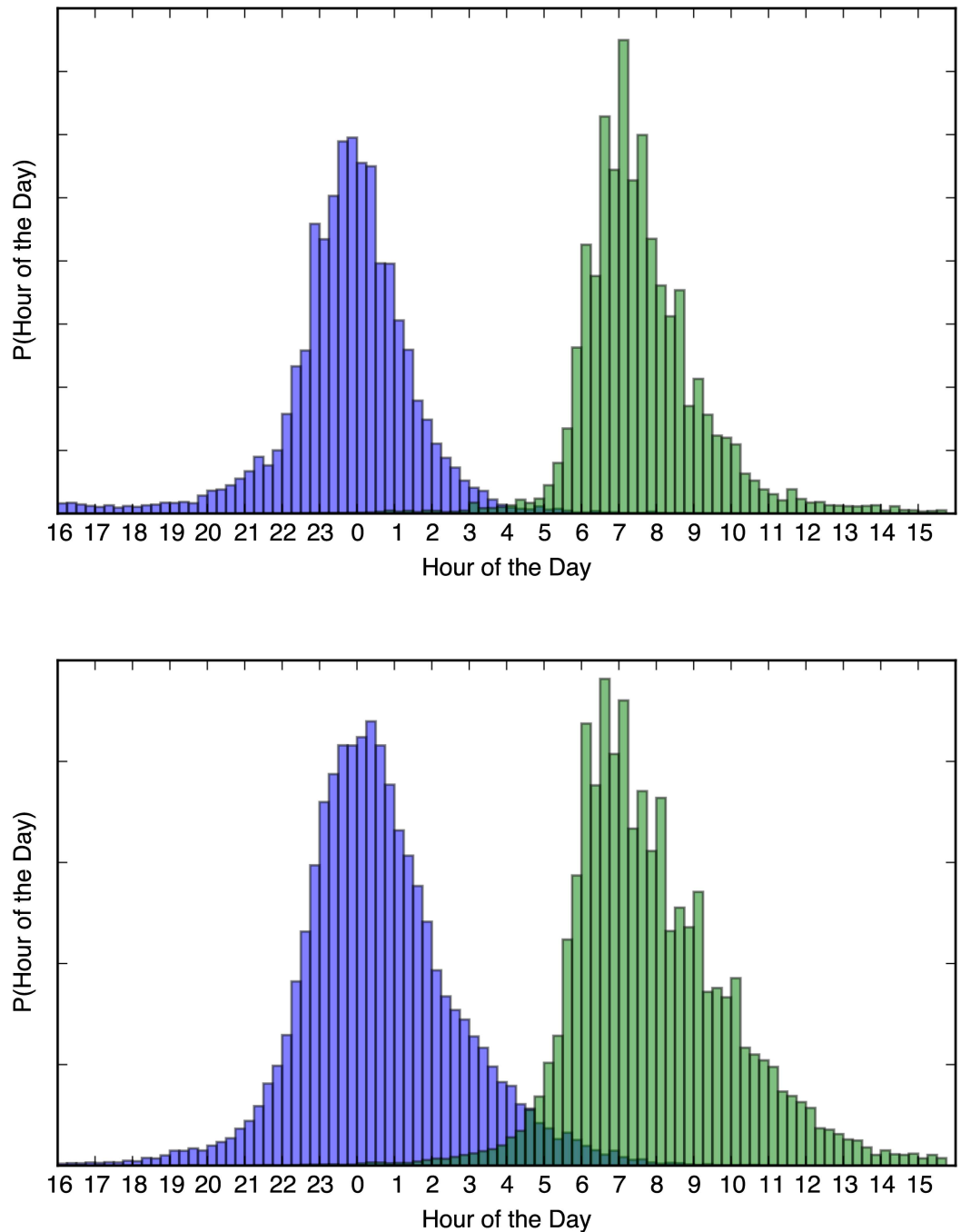
Fig 7. Aggregate Posterior Probability Distributions of t_{sleep} (blue) and t_{awake} (green) (A top, B bottom),
showing what the probability is for the specific population to go to sleep or wake up at the specified time.

the seemingly weak signal of screen events carry significant information of the user status. Our
method has several advantages:

- The method requires only a smartphone and can therefore be deployed without the need for
  special equipment or methods, such as fitness or sleep tracking bands, or sleep diaries.

Table 2. Aggregated sleep and wake times.

	Sleep Time		Wake Time	
	Mean	(Std)	Mean	(Std)
A	23:38	(2h 16m)	7:40	(2h 2m)
B	0:35	(2h 6m)	7:55	(2h 15m)

doi:10.1371/journal.pone.0169901.t002

- The data collection is completely automated, as no action is required from the user in setting up the tracking or remembering to log his/her activity.

- Since the model requires only screen interactions, it is absolutely non-intrusive and privacy-preserving. Although in this work we stored the data on a central server for analysis purposes, the data could remain on the phones and the sleep analysis could in principle be run directly on the phones as well.

- Compared to accelerometer or microphone-based methods, using only screen events is much more battery-efficient.

Although solutions using screen events have been proposed before [8, 21], our model provides a number of key improvements:

- It is more robust to noise such as screen events generated by checking the phone at night.

- Using a Bayesian formulation allows us to provide confidence intervals for the sleep and awake times, instead of point estimates only.

- It does not depend on ad-hoc rules, but it is based on a well-defined statistical formulation.

- It is fitted and verified on a much larger userbase of over 400 users, and a longer time duration (between 2 and 4 weeks).

Demonstrating the feasibility of inferring reasonable sleep patterns from simple event counts opens the way for new exciting research directions. In particular we believe that similar methods can be applied to large datasets of user activity. For example on social network (such as Twitter, Facebook, Meetup, Gowalla) users leave a trace of their activity in the form of messages, posts, likes, etc. Another great example is Call Detail Records, the logging information kept by telecom providers about user calls and SMS. These events could be treated again as a proxy for sleep and wake cycles.

The main drawback of the proposed method is that it requires that users periodically interact with their phones during their wake time. In line with other recent polls (see for example [29–31]), we show that in most cases this does happen, as the population of users analyzed here tend to check their phone from the early morning to the late night when awake. Different populations, however, such as elderly people less accustomed to smartphone usage, may not show similar usage patterns. There is therefore a need for additional work in order to understand how increased sparsity would affect sleep pattern reconstruction.

Conclusions

We have presented a Bayesian model to infer sleep patterns from smartphone interactions, which we have applied to two datasets of more than 400 users in total. We have compared the model output with ground truth from sleep trackers, and we have shown how the model is able to recover the sleep state with a mean accuracy of 0.89 and a mean F1 score of 0.83. Furthermore, we have shown how the model is capable of producing very reasonable individual
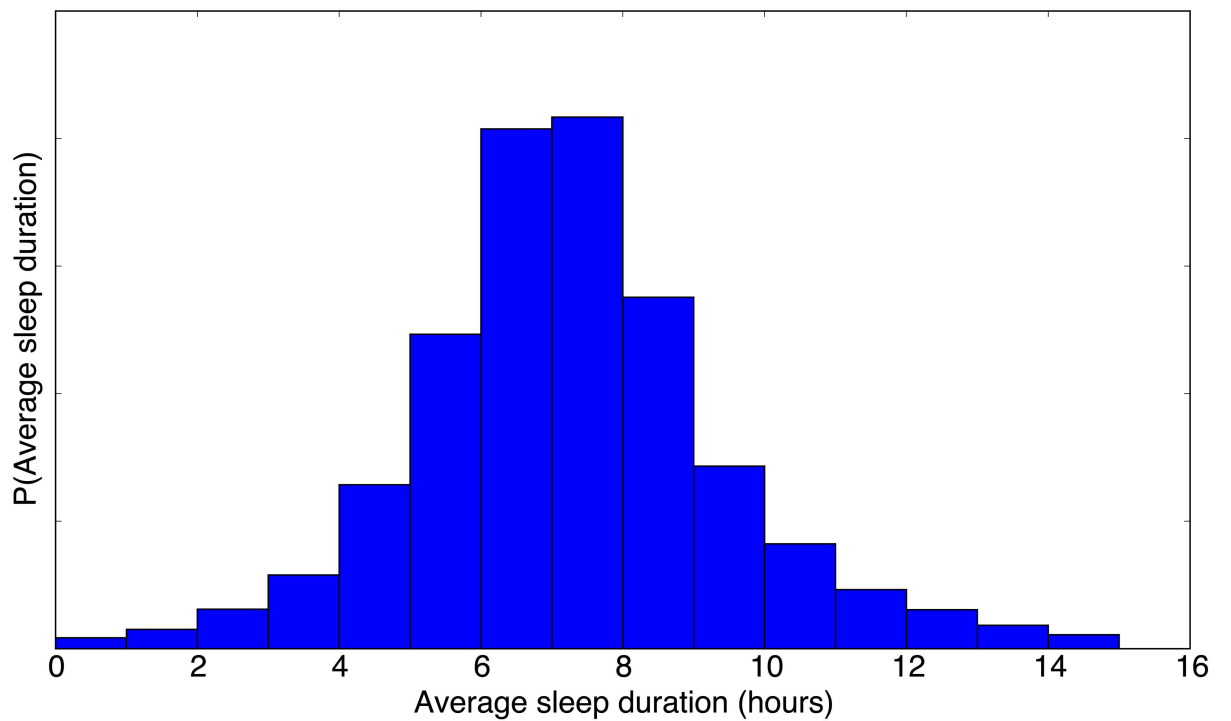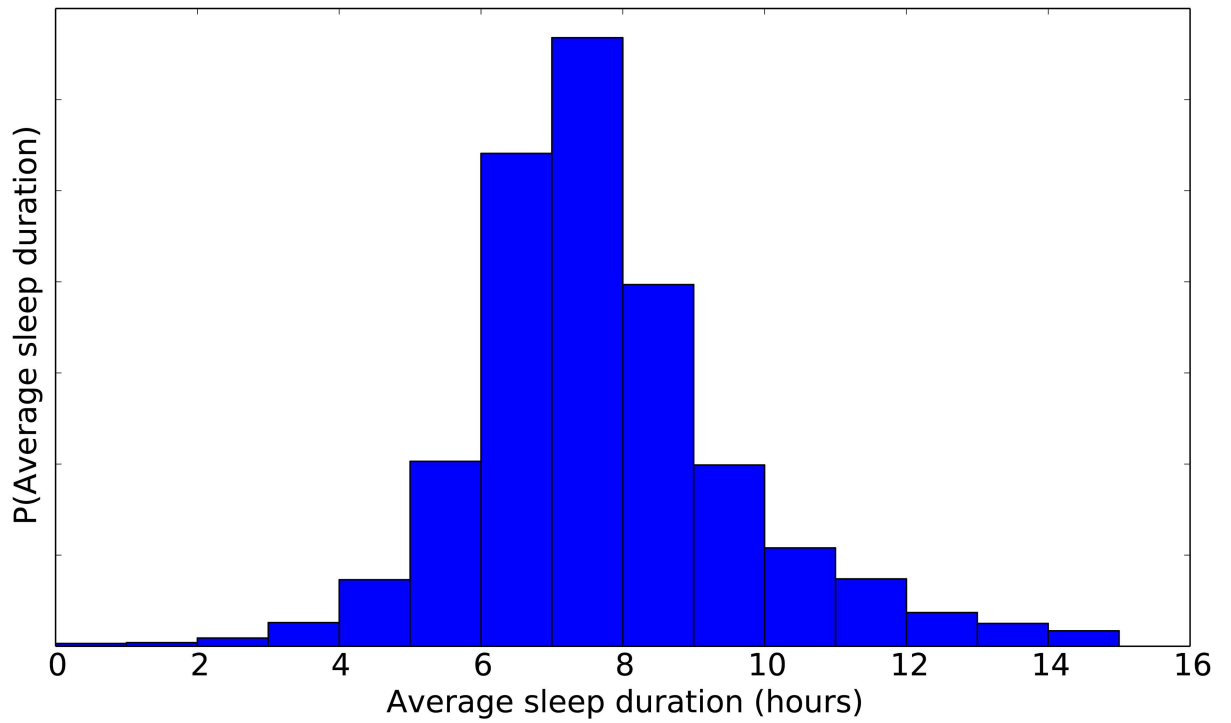
Fig 8. Aggregated Sleep Durations (A top, B bottom), based on the Posterior Probability Functions. This illustrates the probability of the length of a nights sleep within the population within the datasets.
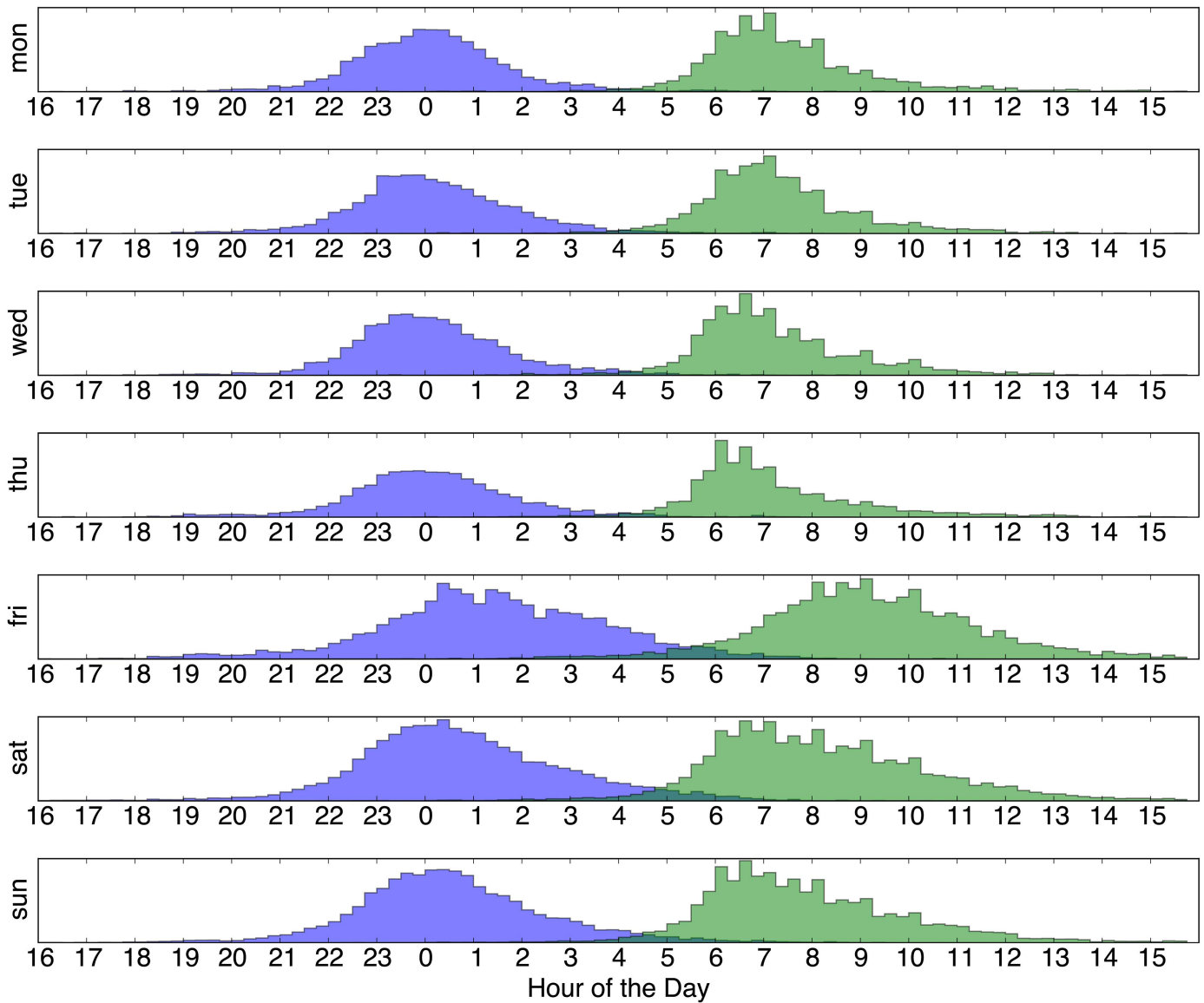
doi:10.1371/journal.pone.0169901.g008

Fig 9. t_{sleep} (blue) and t_{awake} (green) over weekdays for dataset B.

doi:10.1371/journal.pone.0169901.g009

and aggregated sleep patterns. Our method represents a cost-effective, non-intrusive and automatic alternative for inferring sleep patterns, and can pave the way for large-scale studies of sleep rhythms.

Supporting Information

S1 Dataset. Screen interaction events from SensibleDTU. The screen-on events including Unix timestamp and user id.
(ZIP)

S1 Appendix. Poisson vs Power Law Distribution for screen events.
(PDF)

S2 Appendix. Goodness of Fit.
(PDF)

Author Contributions

Conceptualization: AC PB HJ VS JL SL.

Data curation: AC PB.

Investigation: AC PB.

Methodology: AC PB HJ VS JL SL.

Software: AC PB.

Validation: AC PB.

Writing – original draft: AC PB HJ VS JL SL.

Writing – review & editing: AC PB HJ VS JL SL.

References

1. Choe EK, Consolvo S, Watson NF, Kientz JA. Opportunities for Computing Technologies to Support Healthy Sleep Behaviors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI'11. New York, NY, USA: ACM; 2011. p. 3053–3062. Available from: http://doi.acm.org/10.1145/1978942.1979395.

2. Candia J, González MC, Wang P, Schoenharl T, Madey G, Barabási AL. Uncovering individual and collective human dynamics from mobile phone records. Journal of Physics A: Mathematical and Theoretical. 2008; 41(22):1–11. doi: 10.1088/1751-8113/41/22/224015

3. González MC, Hidalgo CA, Barabási AL. Understanding individual human mobility patterns. Nature. 2008; 453(7196):779–782. doi: 10.1038/nature06958 PMID: 18528393

4. Sekara V, Stopczynski A, Lehmann S. The fundamental structures of dynamic social networks. arXiv preprint arXiv:150604704. 2015;.

5. Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, et al. Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(18):7332–7336. doi: 10.1073/pnas.0610245104 PMID: 17456605

6. Zheng Y, Zhang L, Xie X, Ma WY. Mining interesting locations and travel sequences from GPS trajectories. Proceedings of the 18th international conference on World wide web—WWW'09. 2009;(49):791.

7. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. Journal of Medical Internet Research. 2015; 17(7):e175. doi: 10.2196/jmir.4273 PMID: 26180009

8. Abdullah S, Matthews M, Murnane EL, Gay G. Towards Circadian Computing: "Early to Bed and Early to Rise"Makes Some of Us Unhealthy and Sleep Deprived. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014; p. 673–684.

9. Smart Alarm Clock;. https://itunes.apple.com/us/app/smart-alarm-clock-sleep-cycles/id586910133?mt=8.

10. Sleep Cycle;. https://itunes.apple.com/us/app/sleep-cycle-alarm-clock/id320606217?mt=8.

11. SleepBot;. https://play.google.com/store/apps/details?id=com.lslk.sleepbot&hl=en.

12. Sleep as Android;. https://play.google.com/store/apps/details?id=com.urbandroid.sleep&hl=en.

13. Zhang M, Tillman DA, An SA. Global prevalence of sleep deprivation in students and heavy media use. Education and Information Technologies. 2015; p. 1–16.

14. Orzech KM, Grandner MA, Roane BM, Carskadon MA. Digital media use in the 2 h before bedtime is associated with sleep variables in university students. Computers in Human Behavior. 2016; 55:43–50. doi: 10.1016/j.chb.2015.08.049

15. Richter CP. Biological Clocks in Medicine and Psychiatry: Shock-Phase Hypothesis. Proceedings of the National Academy of Sciences of the United States of America. 1960; 46:1506–1530. doi: 10.1073/pnas.46.11.1506 PMID: 16590778

16. Aschoff J. Circadian Rhythms in Man. Science (New York, NY). 1965; 148(3676):1427–1432. doi: 10.1126/science.148.3676.1427

17. Ren Y, Wang C, Yang J, Chen Y. Fine-grained sleep monitoring: Hearing your breathing with smartphones. In: Computer Communications (INFOCOM), 2015 IEEE Conference on. IEEE; 2015. p. 1194–1202.

18. Hao T, Xing G, Zhou G. iSleep: unobtrusive sleep quality monitoring using smartphones. In: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. ACM; 2013. p. 4.

19. Chen Z, Lin M, Chen F, Lane ND, Cardone G, Wang R, et al. Unobtrusive sleep monitoring using smartphones. In: 7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth). IEEE; 2013. p. 145–152. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6563918.

20. Gu W, Yang Z, Shangguan L, Sun W, Jin K, Liu Y. Intelligent Sleep Stage Mining Service with Smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. UbiComp'14. New York, NY, USA: ACM; 2014. p. 649–660. Available from: http://doi.acm.org/10.1145/2632048.2632084.

21. Jayarajah K, Radhakrishnan M, Hoi S, Misra A. Candy Crushing Your Sleep. In: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers. UbiComp/ISWC'15 Adjunct. New York, NY, USA: ACM; 2015. p. 753–762. Available from: http://doi.acm.org/10.1145/2800835.2804393.

22. Hossain HMS, Roy N, Khan MAAH. Sleep Well: A Sound Sleep Monitoring Framework for Community Scaling. In: Proceedings of the 2015 16th IEEE International Conference on Mobile Data Management —Volume 01. MDM'15. Washington, DC, USA: IEEE Computer Society; 2015. p. 44–53. Available from: http://dx.doi.org/10.1109/MDM.2015.42.

23. Sony SmartWear;. http://www.sonymobile.com/global-en/products/smartwear/.

24. Stopczynski A, Sekara V, Sapiezynski P, Cuttone A, Madsen MM, Larsen JE, et al. Measuring large-scale social networks with high resolution. PloS one. 2014; 9(4):e95978. doi: 10.1371/journal.pone.0095978 PMID: 24770359

25. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. vol. 2. Taylor & Francis; 2014.

26. Patil A, Huard D, Fonnesbeck CJ. PyMC: Bayesian stochastic modelling in Python. Journal of statistical software. 2010; 35(4):1. doi: 10.18637/jss.v035.i04 PMID: 21603108

27. Fonnesbeck CJ. PyMC version 3; 2015. Available from: https://github.com/pymc-devs/pymc3.

28. Berg A, Meyer R, Yu J. Deviance information criterion for comparing stochastic volatility models. Journal of Business & Economic Statistics. 2004; 22(1):107–120. doi: 10.1198/073500103288619430

29. Tecmark survey finds average user picks up their smartphone 221 times a day; 2014. http://www.tecmark.co.uk/smartphone-usage-data-uk-2014/.

30. Newport F. Most U.S. Smartphone Owners Check Phone at Least Hourly; 2015. http://www.gallup.com/poll/184046/smartphone-owners-check-phone-least-hourly.aspx.

31. BankOfAmerica. Trends in Consumer Mobility Report; 2015. http://newsroom.bankofamerica.com/files/doc_library/additional/2015_BAC_Trends_in_Consumer_Mobility_Report.pdf.