Modelling Sleep Duration Using Gaussian Processes

Jan Ondras
Department of Computer Science and Technology
University of Cambridge
Cambridge, UK
jo356@cam.ac.uk

Abstract

In this project I cluster users of wearable sensors based on their sleep duration patterns. First, Gaussian Process (GP) models are fitted to users' sleep duration time series and parameters of fitted GP kernels are extracted for each user. These parameters are consequently used to perform hierarchical clustering of users and to interpret the discovered clusters. I compare two kernel types and several choices of parameters used for clustering. My experiments show that the periodic GP kernel is better suited for this problem than the product of Matérn and periodic kernel, and that kernel parameters other than periodicity are not informative for the clustering. My results reveal two main clusters of users who differ primarily in periodicity of their sleeping patterns. Additionally, I provide extensive interpretation in terms of other measures such as bed in and bed out times or number of steps.

1 Introduction

The increasing popularity of wearable devices such as smart-watches allows monitoring of various time series data. Specifically, in this project I focus on users' sleep duration measurements.

Gaussian Process (GP) models have been widely used for time series modelling and forecasting as they can handle sparsely and irregularly sampled data, and most importantly, they provide probabilistic predictions with uncertainty. Application domains include various areas such as meteorology [1], finance [2] or healthcare [3, 4, 5]. In particular, the study [3] employs GPs to model heart rate data and [4] estimates respiratory rates based on the hyperparameters of fitted GP kernels. The work of Pimentel et al. [5] is most similar to mine. They use GP regression to model vital-sign[1] trajectories of patients who are then clustered with the goal of abnormality detection. To perform the clustering they introduce a similarity metric based on the local likelihood of the points in each trajectory.

In contrast to [5], I aim to cluster users using the similarity between fitted kernel parameters as opposed to using the series of predictions made by such a kernel. The main advantage of my approach is that the clustering can be better interpreted since it is performed along several dimensions of kernel parameters which are well interpretable, as compared to the single similarity score employed by [5]. In this project I further compare two GP kernel types with several choices of clustering parameters and provide extensive interpretation of the discovered clusters.

The rest of the report is structured as follows. Section 2 describes the theory of Gaussian Processes relevant for this project, Section 3 outlines the clustering methodology being used and Section 4 provides implementation details. The data analysis and preprocessing is described in Section 5, while the performed experiments and obtained results are presented and analysed in Section 6. Section 7 concludes the report and suggests future research directions.

[1]heart rate, systolic blood pressure, temperature, blood oxygen saturation and respiratory rate

2 Gaussian Process model

Let us consider the Gaussian Process (GP) regression model $y = f(\mathbf{x}) + \epsilon$ which expresses a dependent variable y in terms of an independent variable $\mathbf{x}$, via a latent function $f(\mathbf{x})$ and a noise term $\epsilon \sim \mathcal{N}(0, \sigma^2)$. In my case, the variable $\mathbf{x}$ is one-dimensional and corresponds to time (in days) and y represents sleep duration (in hours). The function f can be interpreted as a random function drawn from a probability distribution over functions, $y = f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, where $m(\mathbf{x})$ is the mean function of the distribution (for simplicity assumed to be zero) and $k(\mathbf{x}, \mathbf{x}')$ is the kernel (or covariance function). Kernels can take a variety of forms as long as they are positive semi-definite. In particular, the kernel $k(\mathbf{x}, \mathbf{x}')$ describes the coupling between two function values $f(\mathbf{x}), f(\mathbf{x}')$ of the independent variable as a function of the distance $r = ||\mathbf{x} - \mathbf{x}'||$ between them.[2] Kernels thus represent our assumptions about the structure of the time series modelled [6] which motivated my choice of the following two types of kernels:

1. Since it is very likely that the sleep duration has a periodic behaviour [7], we can model periodic functions using the periodic kernel defined as
$$k'_P(r|A, \lambda, P) = A^2 \exp\Big(- 2\lambda^{-2} \sin^2(\pi r P^{-1})\Big) \tag{1}$$
where the parameters A, λ and P are amplitude, length-scale and period of the latent function respectively. The amplitude A (or signal variance) can be interpreted as a deviation from the mean and the length-scale λ as describing the behaviour of the signal within a period. To estimate the noise level, the noise term ϵ can be conveniently incorporated into the kernel and its standard deviation treated as another parameter σ.
$$k_P(r|A, \lambda, P, \sigma) = k'_P(r|A, \lambda, P) + \sigma^2 \tag{2}$$
2. It is possible that sleep duration time series also exhibit some long-term variations. To model this we can combine the periodic kernel with a Matérn kernel k_M that can capture such long-term behaviour by the length-scale parameter λ_M. In general, the Matérn kernel is defined as
$$k_M(r|\lambda_M, \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)}\Big(\sqrt{2\nu}\frac{r}{\lambda_M}\Big)^{\nu} K_\nu\Big(\sqrt{2\nu}\frac{r}{\lambda_M}\Big) \tag{3}$$
where Γ is the gamma function and K_ν is the modified Bessel function of the second kind [8]. The parameter ν implies that the corresponding GP kernel has sample paths that are $\lceil \nu \rceil - 1$ times differentiable. I decided to use the Matérn kernel instead of the Squared-Exponential (Matérn with $\nu \to \infty$) since it was argued [9] that such strong smoothness assumptions made by the Squared-Exponential kernel are unrealistic for modelling many physical processes. The most popular choices of ν are $3/2$ and $5/2$ [6]. I chose $\nu = 5/2$ to model a slightly smoother family of functions. In such a case the Eq. (3) can be simplified to
$$k_M(r|\lambda_M, \nu = 5/2) = \Big(1 + \frac{\sqrt{5}r}{\lambda_M} + \frac{5r^2}{3\lambda_M^2}\Big) \exp\Big(- \frac{\sqrt{5}r}{\lambda_M}\Big) \tag{4}$$
Finally, the Matérn and periodic kernels are combined using their product as
$$k_{MP}(r|A, \lambda_M, \lambda, P, \sigma) = k_M(r|\lambda_M, \nu = 5/2) \times k'_P(r|A, \lambda, P) + \sigma^2 \tag{5}$$

The two kernels k_P and k_{MP} are then represented by their parameters $\boldsymbol{\theta}_{k_P} = [A \quad \lambda \quad P \quad \sigma]$ and $\boldsymbol{\theta}_{k_{MP}} = [A \quad \lambda_M \quad \lambda \quad P \quad \sigma]$ respectively. For both kernel types, their parameters $\boldsymbol{\theta}_k$ were selected by maximising the log marginal likelihood $\log p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_k)$ which represents a trade-off between model fit and model complexity. This optimisation process was performed using a search at logarithmic scale with 96 random restarts within the following parameters' bounds: $A \in \big[0.01\sqrt{10}, 10\big]$ hours, $\lambda_M \in [0.5, 1000]$ days, $\lambda \in [0.5, 365]$ days, $P \in [1, 365]$ days, and $\sigma \in \big[0.1, \sqrt{10}\big]$ hours, which were chosen based on the data analysis (Sec. 5).

The predictions about the function values y_* at a test point x_* are given by the Gaussian posterior distribution (conditional on the observed data $\mathbf{x}, \mathbf{y}$) with the mean u and variance s^2 given by
$$\begin{aligned} u &= \mathbb{E}\left[y_*\right] = k(\mathbf{x}, x_*)^T k(\mathbf{x}, \mathbf{x})^{-1}\mathbf{y} \\ s^2 &= \mathrm{Var}\left[y_*\right] = k(x_*, x_*) - k(\mathbf{x}, x_*)^T k(\mathbf{x}, \mathbf{x})^{-1} k(\mathbf{x}, x_*) \end{aligned} \tag{6}$$
When N_d predictions with means $\mathbf{u} \in \mathbb{R}^{N_d}$ and variances $\mathbf{s}^2 \in \mathbb{R}^{N_d}$ are made at a set of test points $\mathbf{x}_* \in \mathbb{R}^{N_d}$, we can define a *mean trajectory* of predictions by $\mathbf{u} \pm \mathbf{s}$.

[2] I focused on stationary kernels that depend only on the distance of two datapoints and not on their absolute values which makes them invariant to translations in the input space.

3 Clustering

For each user i, the set of parameters $\boldsymbol{\theta}_k^{(i)}$ of the fitted GP kernel is extracted which results in the matrix $\boldsymbol{\Theta}_k \in \mathbb{R}^{N_u \times N_k}$ where N_u is the total number of users and N_k is the number of kernel parameters used. The matrix $\boldsymbol{\Theta}_k$ is in turn used to perform hierarchical clustering of users using the Ward variance minimisation algorithm [10] and Euclidean distance metric. Initially, all clusters are singletons. Then, at each step the Ward's method finds a pair of clusters that leads to minimum increase in total within-cluster variance after merging. This increase is calculated as a weighted squared distance between cluster centers. The algorithm terminates when all users are merged into a single cluster.

3.1 Optimal number of clusters

Choosing the optimal number of clusters is challenging and there is no golden method. I chose the variant of the *Elbow method* [11] which tries to find the clustering step where the acceleration of distance growth is the largest. While this method is very flexible, it is unable to discover a single-cluster case and the order of the distances used to calculate the acceleration may not properly reflect the order of merges within one branch of the tree. Therefore, I always confirmed the choice by visual inspection of *dendrogram* (a tree diagram showing the order and distances of merges during the hierarchical clustering).

3.2 Quality of clustering

To asses the quality of the obtained clustering when the ground truth labels are not known, the evaluation is performed using the model itself. For this I used the *Silhouette coefficient* [12] defined as

$$S_c = \frac{1}{|\mathbb{D}|} \sum_{i \in \mathbb{D}} \frac{b_i - a_i}{max(a_i, b_i)} \tag{7}$$

where a_i is the average distance between a sample i and all other points in the same cluster, b_i is the lowest average distance between the sample i and all other points in any other cluster, and $\mathbb{D}$ is the set of all datapoints (users). The Silhouette coefficient is bounded between -1 (incorrect clustering) and 1 (highly dense clustering). Values $S_c \approx 0$ indicate overlapping clusters.

3.3 Similarity between two clusterings

The consensus (similarity) between two clusterings can be evaluated using the *Adjusted Rand Index* [13] R that is bounded between -1 (independent clusterings) and 1 (identical clusterings). This measure has an advantage that it makes no assumptions about the structure of clusters.

3.4 Representation of clusters

Once the clustering is decided upon, we can calculate a refined center of each cluster c from a set $\mathcal{C}$ of found clusters and thus determine the characteristic parameters $\boldsymbol{\theta'}_{k,c}$ for each cluster c as

$$\forall c \in \mathcal{C}. \quad \boldsymbol{\theta'}_{k,c} = \frac{1}{|c|} \sum_{i \in c} \boldsymbol{\theta}_k^{(i)} \tag{8}$$

Also, for every cluster c, we can compute its representative mean trajectory (prototype) $\mathbf{u}_c \pm \mathbf{s}_c$ by calculating a weighted average (according to [14]) as

$$\forall c \in \mathcal{C}. \forall k \in [1 \dots N_d] \qquad \mathbf{u}_{c,k} = \sum_{i \in c} \mathbf{s}_{i,k}^{-2} \mathbf{u}_{i,k} \Big/ \sum_{i \in c} \mathbf{s}_{i,k}^{-2}, \qquad \mathbf{s}_{c,k} = \Big(\sum_{i \in c} \mathbf{s}_{i,k}^{-2} \Big)^{-1/2} \tag{9}$$

where $\mathbf{u}_i \pm \mathbf{s}_i$ is the mean trajectory of predictions of a GP model fitted to user i's data. The average weighted by precisions ensures that less reliable estimates have smaller weights.

4 Implementation

For GP modelling I used the Python framework *scikit-learn* [15] and for hierarchical clustering the toolbox *SciPy* [16]. The presented figures were produced using the plotting library *matplotlib* [17].

5 Dataset

The whole dataset consisted of 10 340 users whose sleep duration was measured on a daily basis during one year, namely, in the range of dates 01/04/2016–31/03/2017 later referred to as the range of days 1–365. This defined the total number of prediction points to be $N_d = 365$ and also motivated my choice of upper and lower bounds used to search for optimal kernel parameters. The dataset also contained more detailed information of bed in and bed out times as well as number of steps per day, and furthermore, for each user the following details were provided: gender, location (San Francisco or London), age, and height. These measures were later used to interpret the clustering in addition to the GP kernel parameters.

5.1 Data analysis and preprocessing

Firstly, the sleep duration measurements of more than 20 hours were eliminated as they were considered to be unrealistic, in fact there were many noisy measurements of more than 24 hours of sleep. Next, I focused only on users with at least 300 measurements per year since for some users there were many missing measurements. These filtering steps resulted in the final set of $N_u = 1264$ users.

As suggested by [3], many natural processes tend to be Gaussian in logarithmic domain and so the log-transform of measurements might be appropriate. To investigate whether this is the case, I compared how many distributions of users' measurements pass the Normality test[3] with and without log-transformed measurements. The obtained results have shown that when the measurements were not log-transformed the distributions of 4.5-times more users passed the Normality test. For this reason, I did not apply the logarithmic transformation of measurements.

Lastly, since the GP model introduced in Sec. 2 assumes zero mean function ($m(\mathbf{x}) = 0$), a prior, an estimate of the global mean sleep duration $\mu_G = 7.26$ hours was subtracted from measurements of all users. The value of μ_G was estimated based on the study [19] by averaging over a week.

6 Experiments, results and analysis

6.1 Choice of kernel type and clustering parameters

Following the theory described in Sections 2 and 3, I fitted a separate GP kernel to whole sleep duration time series of every user and computed the matrices $\mathbf{\Theta}_{k_P}$ and $\mathbf{\Theta}_{k_{MP}}$ (of kernel parameters of all users) for both kernel types k_P (Eq. (2)) and k_{MP} (Eq. (5)) respectively. Using various subsets of parameters from these matrices, I obtained several clusterings and compared them. In all the experiments the optimal number of clusters was determined according to Sec. 3.1 and was found to be the same in all cases, namely, $N_c = 2$ clusters.

The clustering of users using all parameters $\boldsymbol{\theta}_{k_P} = [A \quad \lambda \quad P \quad \sigma]$ of the periodic kernel k_P is shown in Fig. 1, in terms of all pairs of parameters in $\boldsymbol{\theta}_{k_P}$. We can clearly see that only the periodicity P is able to differentiate different kinds of users. Therefore, I further carried out another clustering using only the periodicity parameter which resulted in almost identical clustering, $R = 0.97$ (Sec. 3.3).

For the kernel product k_{MP} (Matérn $\times$ periodic), I performed two clusterings:

1. using all kernel parameters $\boldsymbol{\theta}_{k_{MP}} = [A \quad \lambda_M \quad \lambda \quad P \quad \sigma]$
2. using P and λ_M only, which was motivated by:
   (a) the observation that only the periodicity P is relevant for the clustering based on the kernel k_P,
   (b) and the fact that λ_M is a parameter not present in the periodic kernel k_P and so it might provide another dimension along which the users are separable.

For the second case, the obtained clustering in terms of the parameters P and λ_M is shown in Fig. 2. We can observe that the periodicity is no longer the differentiating factor and this role is taken by the Matérn length-scale parameter. However, this does not seem to provide better clustering as it results in separation along almost only one dimension (λ_M). This suggests that the parameters P and λ_M are competing against each other when the GP kernel is fitted to the sleep duration data.

[3] The test of the null hypothesis that a sample comes from a Normal distribution, based on the D'Agostino and Pearson's test [18] with $\alpha = 10^{-3}$.

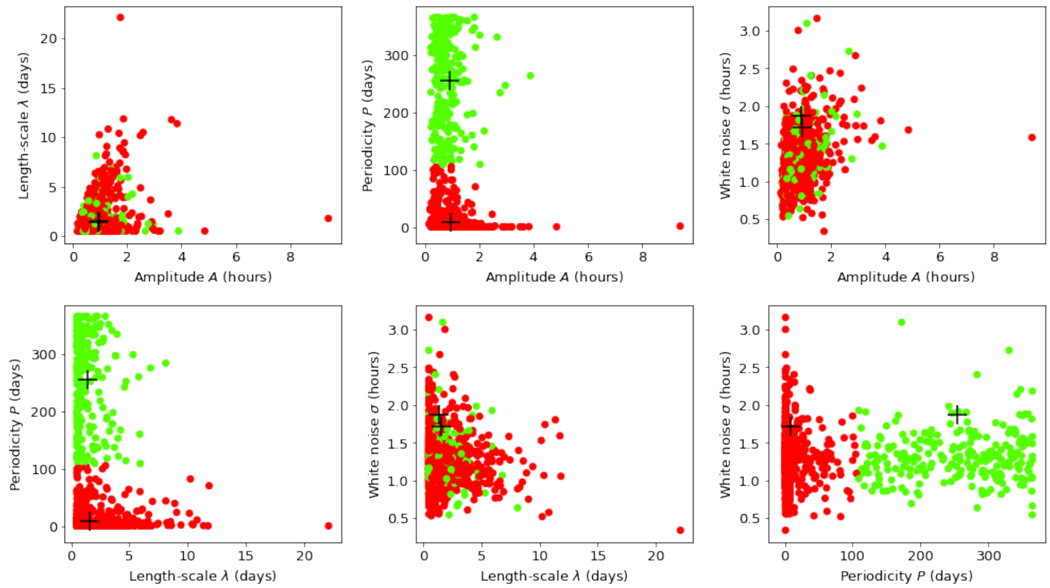

Figure 1: Clustering of 1264 users into two clusters using GP kernel parameters: amplitude A, length-scale λ, periodicity P, and white noise σ. Kernel type: periodic k_P.

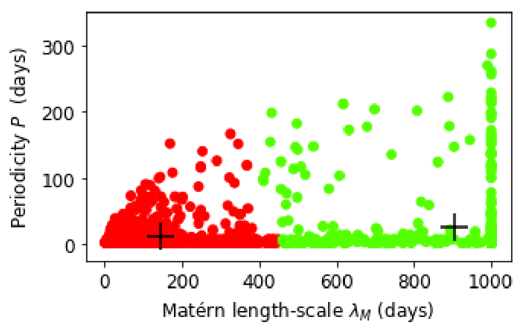

Figure 2: Clustering of 1264 users into two clusters using GP kernel parameters: Matérn length-scale λ_M and periodicity P. Kernel type: product (Matérn $\times$ periodic) k_{MP}.

All the above-mentioned experiments are summarised in Tab. 1 and compared in terms of the Silhouette coefficient (Sec. 3.2) to assess the quality of clusterings. The results based on the Silhouette coefficient confirm the above-described observations that the periodicity P is the only parameter of the periodic kernel that is relevant for clustering and that the kernel product k_{MP} performs worse than the periodic kernel k_P, i.e. the learned kernel parameters $\boldsymbol{\theta}_{k_{MP}}$ lead to worse clustering than $\boldsymbol{\theta}_{k_P}$. Also, as it was anticipated based on the experiments with the periodic kernel, the clustering using only the parameters P and λ_M results in the better-quality clustering than when all 5 parameters of the kernel product are used.

Table 1: Influence of GP kernel type and its parameters on clustering performance.

Kernel	Clustering parameters	Silhouette coefficient (S_c)
Periodic (k_P)	A, λ, P, σ	0.87
	P	0.87
Matérn $\times$ periodic (k_{MP})	$A, \lambda_M, \lambda, P, \sigma$	0.72
	λ_M, P	0.77

It should be noted that in all experiments the parameters were directly used for clustering, since the z-normalisation[4] of parameters has shown to be inappropriate in this case as it gave the same importance to all parameters which resulted in much worse clustering in terms of the Silhouette coefficient and visual inspection of the dendrogram.

Based on the above-mentioned investigations and findings I performed a final clustering of users using the periodicity parameter of the periodic kernel, which is illustrated by a dendrogram in Fig. 3.

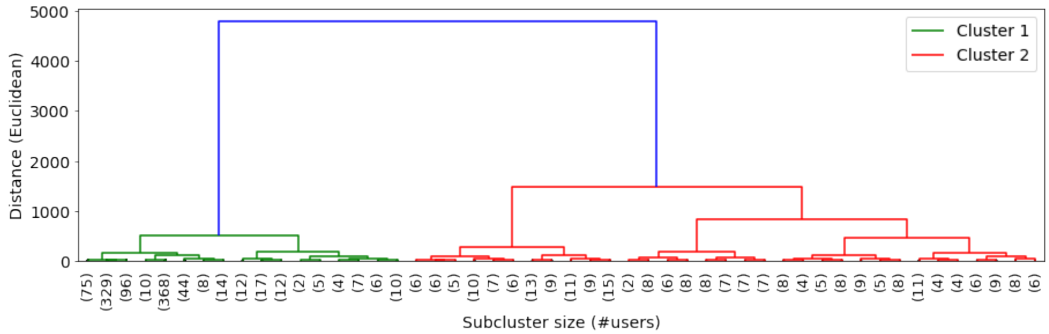

Figure 3: Dendrogram: hierarchical clustering of 1264 users based on the periodicity parameter P of the periodic GP kernel k_P. Optimal number of clusters is 2, which can be also confirmed by the largest gap in the distance metric when 2 clusters are merged into 1.

6.2 Interpretation of clusters

The primary interpretation of the obtained clusters is made in terms of the kernel parameters which according to the findings in Sec. 6.1 reduces to the periodicity parameter P. Using the Eq. (8) I calculated the characteristic sleep duration periodicities of users in Cluster 1 and Cluster 2 to be $P_1' = 8$ days and $P_2' = 250$ days respectively. This shows that users in Cluster 1 tend to have around one-week period of their sleep duration patterns whereas the users from Cluster 2 have much longer periodicity at the scale of several months which seems to be unrealistic and could be interpreted that these users do not have periodic sleep duration patterns at all.

To confirm the different sleep duration patterns between the two clusters, we can visualise the mean trajectories along with the measurements of sleep duration for sample users from both clusters, as shown in Fig. 4.

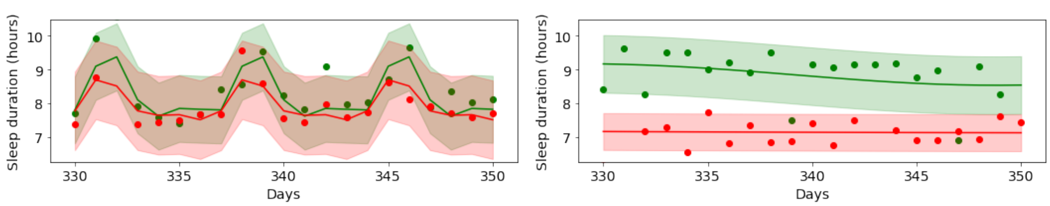

Figure 4: Mean trajectories $\mathbf{u}_i \pm \mathbf{s}_i$ (lines) and measurements (dots) of sleep duration of sample users i from both clusters: Cluster 1 (left) with mean periodicity $P_1' = 8$ days (nearly one-week) and Cluster 2 (right) with $P_2' = 250$ days (hardly any periodicity). Days 330–350 on x-axis correspond to dates 24/02/2017–16/03/2017.

Based on the Eq. (9) I further computed the mean trajectories (prototypes) of each cluster which can be seen in Fig. 5. These prototypes confirm the above-described weekly and longer-term patterns, and moreover, the mean trajectory of Cluster 2 shows that the sleep duration of users from this cluster has its minimum and maximum throughout the year, with the minimum located on day 84 (which corresponds to date 23/06/2016) and maximum on day 286 (date 11/01/2017). This is consistent with the fact that people sleep more during winter and less during summer [20]. We can notice that such a trend is much less visible in the mean trajectory of Cluster 1 which suggests that the sleep duration of users from Cluster 1 is less dependent on yearly seasonality.

[4]Standardisation by removing the mean and scaling to unit variance, along each dimension independently.

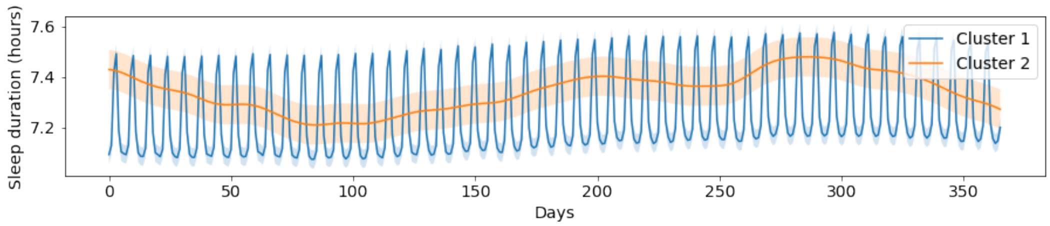

Figure 5: Cluster prototypes: mean trajectories $\mathbf{u}_c \pm \mathbf{s}_c$ representing each cluster c. Days 1–365 correspond to dates 01/04/2016–31/03/2017. Cluster 2 has minimum on day 84 (date 23/06/2016) and maximum on day 286 (date 11/01/2017).

The interpretation of the clusters can be further enriched using other available information (Sec. 5) about users and using aggregate measures. The former is summarised in Tab. 2 and the latter by Fig. 6 where the sleep duration, bed in and bed out times, and the number of steps are aggregated by days in week.

Table 2: Characteristics of 1264 users clustered into two clusters.

		Cluster 1	Cluster 2
Size (# users)		1019	245
Gender:	Males	50%	50%
	Females	51%	49%
Location:	San Francisco	52%	48%
	London	49%	51%
Mean age (years)		45.6	47.2
Mean height (m)		1.74	1.72

As it can be seen from Tab. 2 the two clusters are well balanced in terms of the gender and location whereas slightly younger and taller users tend to be in Cluster 1. The distributions of gender and location were normalised by cluster size.

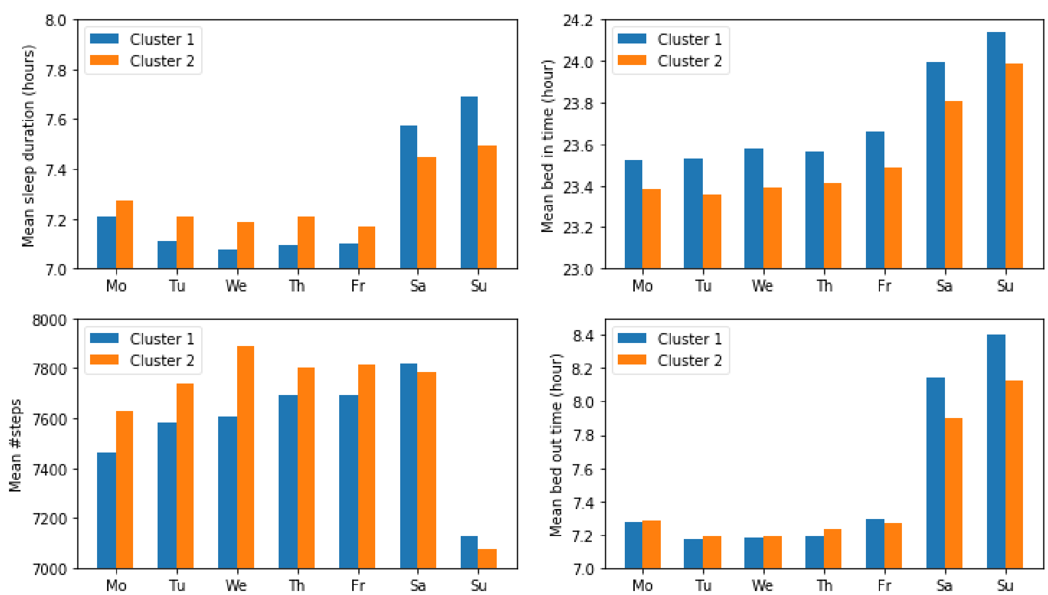

Figure 6: Characteristics of 1264 users clustered into two clusters, in terms of the mean sleep duration, bed in time, bed out time and the number of steps for each day in week.

The results depicted in Fig. 6 can be summarised as follows.

- Characteristics of users in Cluster 1 when compared to Cluster 2:
  - Shorter sleep duration on weekdays but longer during weekends.
  - Later bed in time on all days.
  - Similar bed out times on weekdays but later during weekends.
  - Lower step counts on weekdays but similar during weekends.
- Users in Cluster 2 tend to walk considerably more on Wednesdays as compared to other days (which may reflect some regular activity such as an exercise session).
- Characteristics of users from both clusters:
  - Sleep duration tends to decrease from Monday to Friday which is mostly caused by increasing bed in times and not by decreasing bed out times.
  - Sleep duration, bed in, and bed out times increase during weekends.
  - There is a significant drop in step count on Sunday.

To ensure that the observed weekly trends are not due to different age or height distributions between clusters but due to the clustering performed, I downsampled both clusters to align their distributions of age and height (in terms of means and standard deviations). Namely, 17 users were removed from Cluster 2 and 180 from Cluster 1. The difference in figures was negligible which confirms that the clustering based on the periodicity parameter is able to distinguish the two above-described groups of users.

7 Conclusion

In this project I have used Gaussian Processes to model sleep duration patterns of 1264 users of wearable devices and clustered them based on the fitted kernel parameters. I have found that for this task the periodic kernel performs better than the product of Matérn and periodic kernel, and that the kernel parameters other than periodicity do not bring any additional information into clustering.

My results have shown that there are two main clusters of users. The 1019 users from the first cluster show near one-week periodicity in their sleep duration patterns, and when compared to the other cluster their sleep duration is less dependent on yearly seasonality, they have slightly lower mean age and higher mean height, tend to sleep less during weekdays and catch up during weekends, go to bed later on all days, wake up later during weekends, and walk less on weekdays. The second cluster consists of 245 users whose sleep duration periodicity is at the scale of several months (suggesting no periodicity at all), and with other characteristics opposite to the first cluster. For some reason, users from the second cluster walk on Wednesdays considerably more than on other days.

Further work in this area could use other time series data such as step counts or heart rate and fit GP models along another dimension. One possibility would be to fit the GP models along each dimension independently. Alternatively, a multi-task GP framework [21] could be used to model multiple correlated time series simultaneously. In both cases, this might provide richer space of parameters relevant for clustering.

Acknowledgments

I am grateful to Luca Maria Aiello and Daniele Quercia from Nokia Bell Labs for their guidance and access to the dataset.

References

[1] Niya Chen, Zheng Qian, Ian T Nabney, and Xiaofeng Meng. Wind power forecasts using gaussian processes and numerical weather prediction. *IEEE Transactions on Power Systems*, 29(2):656–665, 2014.

[2] Stephen J Taylor. *Modelling financial time series*. world scientific, 2008.

[3] Oliver Stegle, Sebastian V Fallert, David JC MacKay, and Søren Brage. Gaussian process robust regression for noisy heart rate data. *IEEE Transactions on Biomedical Engineering*, 55(9):2143–2151, 2008.

[4] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. Probabilistic estimation of respiratory rate using gaussian processes. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 2902–2905. IEEE, 2013.

[5] M Pimentel, D Clifton, Lei Clifton, and Lionel Tarassenko. Modelling patient time-series data from electronic health records using gaussian processes. In *Advances in Neural Information Processing Systems: Workshop on Machine Learning for Clinical Data Analysis*, pages 1–4, 2013.

[6] Carl Edward Rasmussen and Christopher KI Williams. Gaussian processes for machine learning. 2006. *The MIT Press, Cambridge, MA, USA*, 38:715–719, 2006.

[7] Shi Ngar Wong, Mark Halaki, and Chin Moi Chow. The periodicity of sleep duration–an infradian rhythm in spontaneous living. *Nature and science of sleep*, 5:1, 2013.

[8] Milton Abramowitz and Irene A Stegun. With formulas, graphs, and mathematical tables. *National Bureau of Standards Applied Mathematics Series e*, 55:953, 1965.

[9] Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

[10] Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

[11] J Hees. Scipy hierarchical clustering and dendrogram tutorial. *Joern's Blog, Web blog post*, 26, 2015. [Online; accessed 19/02/2018].

[12] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[13] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

[14] Marco AF Pimentel, Peter H Charlton, and David A Clifton. Probabilistic estimation of respiratory rate from wearable sensors. In *Wearable electronics sensors*, pages 241–262. Springer, 2015.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[16] Eric Jones, Travis Oliphant, and Pearu Peterson. SciPy: open source scientific tools for Python. 2014. http://www.scipy.org/ [Online; accessed 10/01/2018].

[17] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[18] Ralph D'Agostino and Egon S Pearson. Tests for departure from normality. *Biometrika*, 60(3):613–622, 1973.

[19] Alan T Piper. Sleep duration and life satisfaction. *International Review of Economics*, 63(4):305–325, 2016.

[20] M Kohsaka, N Fukuda, K Honma, S Honma, and N Morita. Seasonality in human sleep. *Experientia*, 48(3):231–233, 1992.

[21] Robert Dürichen, Marco AF Pimentel, Lei Clifton, Achim Schweikard, and David A Clifton. Multitask gaussian processes for multivariate physiological time-series analysis. *IEEE Transactions on Biomedical Engineering*, 62(1):314–322, 2015.