

# Built Environment Characteristics, Daily Travel, and Biometric Readings: Creation of an Experimental Tool based on a Smartwatch Platform

Robin C.O. Palmberg<sup>a</sup>, Yusak O. Susilo<sup>a</sup>, Győző Gidófalvi<sup>b</sup>, Fatemeh Naqavi<sup>a</sup>

a. Department of Urban Planning and Environment, Division of Systems Analysis and Economics, KTH Royal Institute of Technology, Teknikringen 10A, SE-114 28 Stockholm, Sweden

b. Department of Urban Planning and Environment, Division of Geoinformatics, KTH Royal Institute of Technology, Teknikringen 10A, SE-114 28 Stockholm, Sweden

## Disclaimer

This work is currently not under review but is intended to be submitted to the “International Conference on Transport Survey Methods” once the call for submissions is opened.

## Abstract

The utilisation of travel surveys can uncover layers of information regarding travel behaviour, travel needs, and more. The collected information is utilised to make strategic planning choices when reorganising or planning new built environments. Over the years, the methods for conducting travel surveys have changed from manual interviews and paper forms to automated travel diaries which are monitoring the trips made by the survey participants. With the fast progression of technological advancements, new possibilities for operationalising said types of automated travel diaries can be changed from utilising mobile devices to wearable devices. Wearable devices are often equipped with sensors which can collect continuous biometric data from sources which are not reachable from standard mobile devices such as smartphones. The biometric data that can be collected through wearable devices ranging from heart rate and blood pressure to temperature and perspiration, given the proper sensors. This advancement opens for new possible layers of information in the collection of travel data. Such biometric data can be used to derive psychophysiological conditions related to cognitive load, which can uncover more in-depth knowledge regarding stress and emotions, given the right variables and sample rate. This paper aims to explore the possibilities in terms of data analysis on a data set collected through a software combining traditional travel survey data, such as position and time, with biometric data, in this case; heartrate, to gain knowledge of the implications of such collected data. The knowledge about the implications of

spatial configurations can be used in the planning phase of new areas, in order to create more accessible environments, as the information could be used to make neutral, or even encouraging, environments for travellers.

Keywords: Built environment, position data, biometric data, automated data collection, urban planning, traveller needs, traveller behaviour

# 1 Introduction

Travel surveys have been used for decades to observe patterns, locations, and choices made by travellers, in an attempt to understand their needs and behaviour (Axhausen et al., 2002; Chen et al., 2016). By utilising this information, it is possible to make strategic planning decisions and monitor the outcomes of such decision realisations. As we are striving for an equal society, it is crucial to understand why and how inhabitants are affected by their surroundings in order to create more accessible areas. There are currently several ways of conducting travel surveys, including traditional pen and paper surveys (Axhausen et al., 2002), interviews as well as semi-automated collection methods utilising mobile technologies to create travel diaries for the travellers (Allström et al., 2017; “MEILI - travel diary,” 2017, “TRavelVU,” 2016; Adrian C. Prelipcean et al., 2018a; Prelipcean et al., 2014). In the strive to acquire more reliable and trustworthy data, these methods that are verging on automatic data collection are developed to use machine learning methods to predict, analyse and label the collected data. The progression has been rapid, and it is now possible to not only collect the data about where the traveller is travelling but also possible to predict or conclude the chosen travel mode using a fusion of several sensors such as gyroscope and accelerometer (Ferrer and Ruiz, 2014). This progression in collection methodology makes it possible for the survey participant to travel as usual and then review their collected data post-collection, with the possibility to either acknowledge and approve the collected data or change it where there might be incorrect data. The further the development of such methods goes, the more independently the methods will collect the data.

However, the trend of utilising additional data sources to predict additional data, such as the accelerometer and gyroscope data, which is predicting travel modes utilising machine learning methods, opens for the possibility to collect additional biometric data regarding the survey participant. With the currently available technology, it is not only possible but also easily accessible, to obtain biometric data such as heart rate using wearable devices that resembles watches (Apple Inc, 2019a; Benedetto et al., 2018; Empatica, 2019; Fitbit, 2019a; Garmin Ltd., 2019; Gorny et al., 2017; Huawei Technologies Co., 2019; Samsung, 2019a, 2019b; Wallen et al., 2016). By adding the collection of such biometric data, it is possible to extract more information which might relate to psychophysiological conditions, such as cognitive load and stress, from its biometric sensors (Hosseini and Khalilzadeh, 2010; Jerčić

et al., 2018; Lee et al., 2004; Setz et al., 2010; Sharma and Gedeon, 2013; Wijsman et al., 2013; Zhai and Barreto, 2006). The addition of the possibility to derive such information opens up a new perspective on how travel surveys can be enriched, analysed, and deployed.

To explore this new field, we have been experimenting and designing a method which combines biometric data and traditional travel information related to positioning and is currently in the development stage of software that implements the said method (Palmberg et al., 2019a, 2019b). The said method has previously been trialled and showed that it is possible to collect the required data. This new data collection method opens for the possibility to analyse the type of data collected through said software to see what type of information can be derived from such data collections.

## 1.1 Overview

This paper will cover the data analysis, of a data set collected through the use of a proof-of-concept stage software in order to explore the possibilities to uncover how travellers are affected by the built environment, by comparing biometric data deviations at specific locations, data which has been collected automatically using positioning and biometric sensors in combination. The background section aims to cover related work and what the result of that work could implicate for this paper's conclusion. The methodology will be divided into two parts; 1) one part which covers how and what data is collected and the processes revolving the collection of the data set for this particular paper and 2) another part which covers analysis methods explored in the process of analysing said data set. The results will be presented in the section following the methodology, and a discussion evaluates the results based on the recent work presented in the background. Finally, the conclusion section will conclude the trial and point out key findings from work presented in the paper, focusing on the implications of utilising biometric data as a feature for travel surveys which result is aimed at urban and transport planners in general.

## 2 Background

This section aims to cover related articles and reports in order to create a background for the paper. The chosen references are related to the collection and analysis of data through mobile and wearable devices in order to show the possibilities the recent discoveries have granted the research field of the authors of this paper, which is transport planning as well as geoinformatics and urban planning. This section also aims to present the background for the data analysis methods used for this paper.

## 2.1 Data collection

In order to create a background for the intended analysis methods which will be described and utilised in this paper, this section aims to explain how travel diaries function and what biometric data could add to enrich the result of the analysed data. For more information on how these ubiquitous datasets, in general, can be used to analyse and model travel choices and behaviours, please see Chen et al. (2016), Li et al. (2016) and Rashidi et al. (2017), for examples.

### 2.1.1 Collection of biometric data

Recent technological advancement has led to the development of biometric sensors being implemented in consumer-grade hardware (Apple Inc, 2019a; Empatica, 2019; Fitbit, 2019a; Huawei Technologies Co., 2019; Samsung, 2019a, 2019b). This development means that the collection of biometric data becomes available at even more substantial quantities than before since it is possible to create software implementing biometric data collection for widely available platforms, such as Android Wear and watchOS (Apple Inc, 2019b; Google LLC, 2019a). Much like automated travel diaries for mobile phones, software for these widely available platforms make it possible to reach device users in a more effective manner than before, since it requires less activity from both the device user to participate and for the researcher to distribute software and hardware as the user can utilise their own hardware and the researcher can distribute their software through operating system dependant marketplaces (Apple Inc, 2019c; Google LLC, 2019b). Recent research shows proof that it is possible to deduce mental conditions, such as cognitive load, through the use of such types of hardware (Jerčić et al., 2018). By combining several sensors, the different data streams can be used to confirm and imply these conditions. Modern consumer-grade hardware in the form of wearables often includes several types of sensors to deduce activities of the user (Fitbit, 2019b; Garmin Ltd., 2019; Silbert, 2019). Such sensors include, but are not limited to, accelerometer and gyroscopic sensors (from here on referred to as “motion sensors”) for motion detection, ECG and/or PPG sensors (from here on referred to as “biometric sensors”) for heart rate and/or blood pressure as well as GPS and GLONASS (from here on referred to as “position sensors”) for positioning (American Heart Association, 2015; Ferrer and Ruiz, 2014; Healthy Doc, 2018; Jovanov et al., 2001; Li et al., 2015; Murakami and Wagner, 1999; Numenta, 2014; Prelipcean et al., 2014; Simjanoska et al., 2018). The motion sensors collect information about how the device is put in motion, which then can translate to the deduction of several activities such as walking, running, climbing stairs and more (Fitbit, 2019b; Garmin Ltd., 2019). This information in combination with the biometric sensors, which often collect heart rate, can help in confirming those activities by comparing the level of activity with the current heart rate (Gorny et al., 2017). These sensors can also be used in combination

with positioning sensors for popular activity tracking apps in order to analyse and collect data about progress, display maps over past activities and more (ASICS Digital Inc, 2019).

## 2.1.2 Automated travel diaries

Automated travel diaries are an efficient way to collect data for analysis of travel behaviour and needs of travellers in urban areas (Adrian C. Prelipcean et al., 2018a). The development of such diaries is striving to become as automated as possible, both when it comes to the collection of trip legs, but also regarding mode selection. Much like the activity applications described in the last paragraph, the travel diaries can utilise motion sensors in combination with the positioning sensors to deduct travel modes as each mode has a difference in the movement signature (Bohte and Maat, 2009; Gonzalez et al., 2010; Reddy et al., 2010, 2008). This type of deduction is performed by comparing previous data with the new input in order to find the most likely answer to the question of mode choice. With the same methodology, it is possible to find deviations in the movement patterns. Traditional travel diaries often incorporate the recording of such factors, as well as metadata, which could be the reason for travelling to the destination and more, entirely manually (Axhausen et al., 2002). Prelipcean et al. (2018b; 2018; 2017) and Wang et al. (2018), for example, has provided inventories of various operational travel diary tools that have been deployed until recently. For a summary on issues that still need to be further developed to create fully automated travel diaries, Prelipcean and Yamamoto (2018), Gadziński (2018), Stopher et al. (2018), and Wang et al. (2018) can be referred.

The related work presented in 2.1 has shown the possibilities of collecting a wide range of data types through simple means available to the general public, and therefore possible for researchers to use in extensive quantity data collections. Next section will describe how the collected data can be analysed through different data-driven methods.

## 2.2 Analysis methods

Machine learning has been a relevant topic in the field of data analysis for years (Ho Yu, 1977). It simplifies the process of applying statistical methods to data sets in order to capture relevant results from it. Instead of programming complex rulesets for each specific data set, machine learning provides the possibility to reuse more general algorithms for many data sets, without having to program for specific scenarios. There are many methods to choose from, and just as in traditional statistics, the different methods will uncover different results, depending on what has been explored. In this section, clustering and regression methods will be described and compared as a basis for the methodology of the data analysis in this paper.

## 2.2.1 Partition clustering

As the focus of this paper is to analyse biometric and positioning data in order to find how areas and locations might affect travellers, a strategy for finding out if areas within the data set are affecting the participant at all is partition clustering. Partition clustering is general methods which use all dimension of the data set to find patterns within which the samples can be organised.

A common partition clustering method is called “K-means clustering”, which is using all the dimensions of the data set in order to find the “distance” between each data point and the  $K$  number of reference points, taking all dimensions into consideration (MacQueen, 1967). The reference points are at first put at relatively random positions in which their distance to the data points are measured, and they are then moved between each run of the algorithm until they are put in such a position in which they are the closest to “their” cluster of data. This type of clustering method is useful for when one is trying to find which data points are related to each other and which are not without any prior knowledge of the data set. When the method has compared the distances and moved the reference points to locations from which they are no longer possible to move to a more optimal position, the analysis is done, and the points of data are divided according to which reference point they are closest to. The formula of k-means clustering used is as follow:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (\text{eq.1})$$

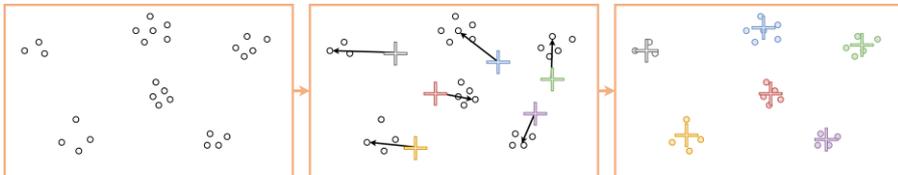


Figure 1 - The process of K-means clustering

In order to find  $K$ , there are several methods to analyse the data set (Kassambara, 2018). One example is the “elbow method”, in which one runs K-means clustering on the data set iteratively with the integer  $K$  changing incrementally after each final round between two values, for example, 1 to 10. For each  $K$  one calculates the sum of squared errors (SSE) and projects the result of a graph, with the value of  $K$  as the X-axis and the SSE as the Y-axis. At one point, an apparent elbow in the line should be visible, indicating that the  $K$  value of that point is the optimal  $K$  for the K-means clustering.

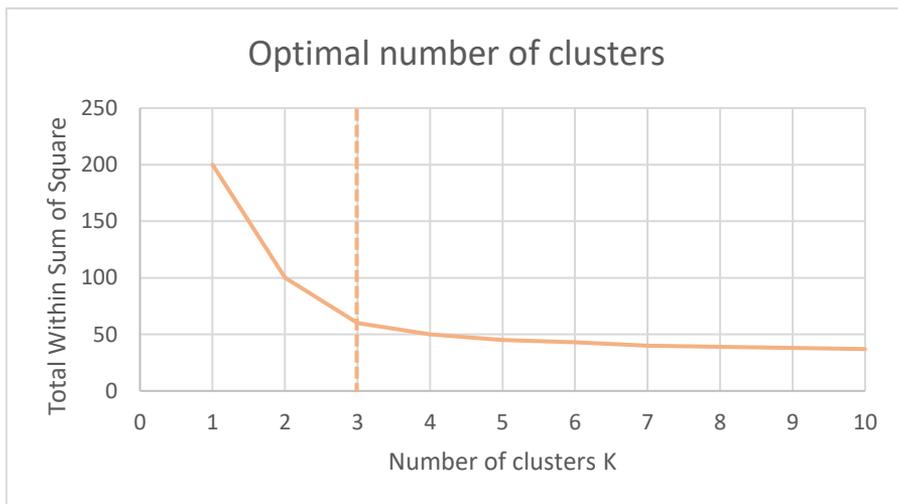


Figure 2 - A diagram depicting a typical elbow found using the elbow method

However, the elbow method can sometimes be ambiguous as it might be hard to distinguish at what value of  $K$  the elbow is. A less ambiguous method is the Average Silhouette Evaluation (ASE) method. The algorithm is similar to that of the elbow method but deviates in the sense that instead of calculating the SSE of each value  $K$ , it measures the quality of the clustering by iterating and finding the optimal  $K$  when the average silhouette is maximised within the set range of possible  $K$  (Kaufman and Rousseeuw, 1990).

K-means clustering is a versatile algorithm as it is unsupervised and therefore can work directly with the intended data set. It also accounts for each dimension equally, meaning that it does not put extra weight on certain features, such as position or time. This can, however, be a drawback when working with spatial data, such as the one described in this paper. Therefore, other methods will be described in the following paragraphs.

## 2.2.2 Spatial-based clustering

When working with spatial data, it might be beneficial to emphasize the spatial dimension of the data. There are several methods to cluster data with regards to spatial dimensions such as density, spatial restrictions and time intervals. As the focus of this paper is to find enclosed areas which can be compared, a suitable method for clustering the data is density-based clustering. Focusing on the spatial density of collected data will output clusters where large amounts of data have been collected in an enclosed area. These areas could then be further analysed in regard to biometric data in order to uncover where and how travellers are affected within that area. By comparing similar areas with similar effects using data-driven methods,

it should be possible find patterns among said areas related to the specific spatial configurations and features of those locations.

Density-based spatial clustering of applications with noise (DBSCAN) can be used to cluster density-based spatial data and filter out outliers within the data set as noise (Ester et al., 1996). Instead of using reference points with a dynamic radius to find a mean between data points, as K-means clustering does, DBSCAN utilises a set maximum radius and minimum point density. The algorithm is applied to each data point to measure how many other data points are located within the maximum radius of the point. If more than the minimum amount of points is located within the radius, that point is labelled as a core point which belongs to a cluster. If less than the minimum amount of points is within the radius, but at least one point within the radius is a core point, that point is a border point, which means that the analysed point is located at the border of, and belongs to, the cluster. If no points are found within the radius or none of the found points are core points, the analysed point is an outlier and is considered as noise.

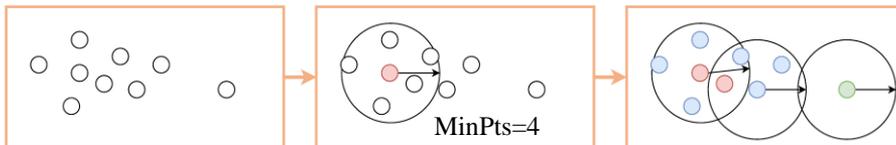


Figure 3 - The iterative process of DBSCAN; red points represents core points; blue points represent border points and green points are outliers

As with the K-means clustering, DBSCAN needs some variables to function. In K-means clustering, the number of clusters needs to be defined, but in DBSCAN, the minimum number of points and maximum radius needs to be defined. This can be done through the use of the elbow method, which is less ambiguous for DBSCAN as it can use decimal values rather than integers, which K-means clustering requires.

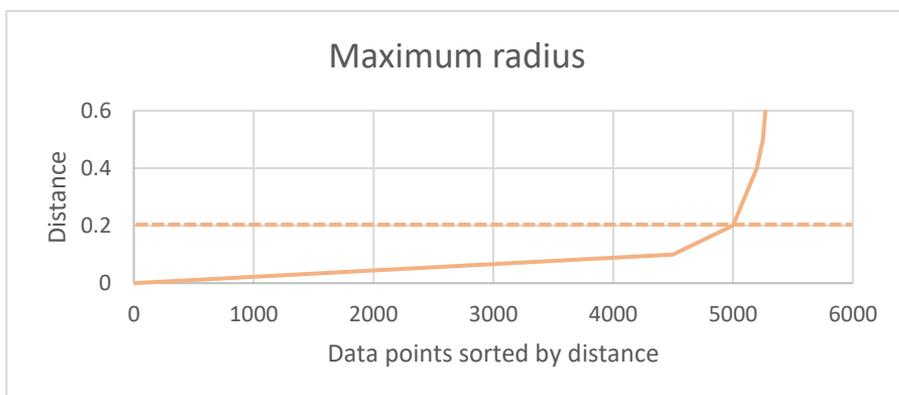


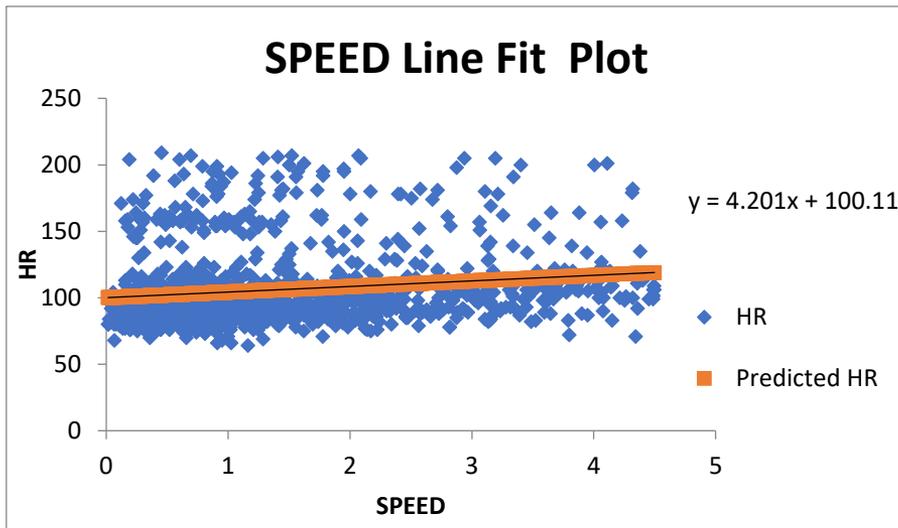
Figure 4 - A diagram depicting a typical elbow found using the elbow method

This method is more efficient than K-means clustering in the sense that it can cluster data in asymmetric forms, in comparison to K-means clustering which creates a radius from the reference point in which all cluster points need to be located. It also allows for outliers to be considered as noise, as opposed to K-means clustering, which uses all data points, no matter if they are outliers or not. However, it does not uncover anything related to correlation as it only clusters the data for further analysis. To understand how the collected data correlates, other methods will have to be used, which are described in the following section.

### 2.2.3 Regression

The previous sections have presented and described clustering methods which do not put emphasis or weight on any specific dimension (K-means clustering) or only focuses on positioning dimensions (DBSCAN). Said methods are applicable for finding patterns in the data in order to group a dataset into several clusters, but there is no information regarding the correlation between dimensions that can be found using said methods. However, there are other methods that are capable of comparing dimensions in order to find the correlation between them.

One method to compare data in order to find patterns is regression analysis. The main difference between clustering and regression is that instead of finding a mean or a density-based body, the output of a regression analysis is the variable for an equation describing the correlation between the data dimensions that are being compared (Galton, 1886; Lai et al., 1979; Rencher and Christensen, 2012; Zhang, 2004). The simplest form of regression is called linear regression, in which the goal is to find the variables for a line that aligns with the trend of the comparisons between the data dimensions. One of the dimensions is dependent on the other, which can be described with the equation found in Figure 5.



$$Y = aX + b$$

*Figure 5 - A diagram of a line function deducted from using linear regression*

In this equation,  $Y$  is the variable that depends on  $X$ , and the aim of utilising the linear regression is to find the constant  $b$  and the regression coefficient  $a$ . When the constant and regression coefficient has been found, the equation can be used to 1) find the probability that there is a correlation between the dependent and the independent variable, or 2) predict  $Y$  based on  $X$ , opening up for the possibility to predict dependent variables based on independent data. This prediction is useful when analysing travel behaviour and travel needs as it can uncover a general understanding of how a particular factor in travelling is affecting the travellers. There are many types of regression analysis that can be formed, but for the sake of the work proposed in this paper, only linear and multivariate regression will be covered. Next, the methodology will explain how the data set was collected, how the data was refined, and how it was analysed.

### 3 Methodology

Based on the background in the previous section, this section aims to explain how the data for the analysis has been collected and what methods for data collection have been used with arguments for why. As mentioned in the introduction, this paper covers the trial of analysis of a data collection tool developed by the first author called MERGEN (PalMBERG et al., 2019a). The trial in this paper covers the collection of data from a small set of participants and emphasises on the analysis methods of said data, and what said data could reveal about a traveller. This is the first step in a series of trials in order to uncover the possibilities to combine and collect position and biometric data using the tool.

#### 3.1 Data collection

The framework for data collection used in this trial is called MERGEN and was designed based on previously available semi-automated methods of travel data collection, such as those described in section 1 and 2, augmented with the addition of biometric data. The framework can be used with a range of devices, but the version used in this trial is a software, based on a wearable specific version of Android, called Wear OS, and utilises consumer-grade hardware, such as smartwatches, which include biometric sensors. In this particular case, a smartwatch with a PPG sensor designed for the collection of heart rate was used. The framework combines the collection of position data and adds the dimension of heart rate to the collected data set, as visualised in Figure 6.

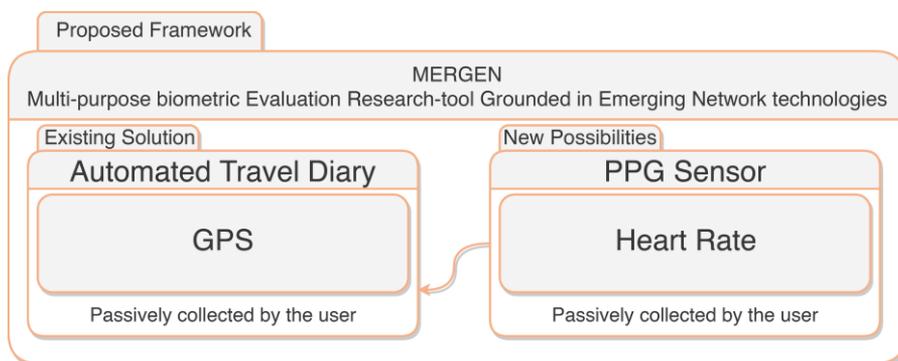


Figure 6 - Schematic of the framework MERGEN

The software is intended to be useable by any Wear OS-compatible device and is therefore designed to utilise the built-in functions for the collection of position and biometric data. This opens for the possibility to choose among a wide range of devices, with several variations of combinations of sensors. As the software had been developed and tested with the use of a Huawei Watch 2 4G, the same model of the

smartwatch was used in this trial, as not to add any additional factors which could disturb the outcome. The reason as to why this specific watch was used in the development process was because of the versatile range of sensors, shown in Table 1.

*Table 1 - The technological specifications of the utilised hardware*

<i>Specifications</i>	<i>Descriptions</i>
<i>Operating System</i>	Wear OS 2
<i>Sensors</i>	PPG, GPS + Glonass, 6-axis A+G Sensor, 3-axis Compass, Barometer, Capacitive Sensor, Light Sensor
<i>Communication</i>	4G LTE, BLE, NFC (RFID), UMTS, TD-SCDMA, GSM
<i>Screen</i>	1,2-inch Circular AMOLED with 326 PPI
<i>Battery</i>	410 mAh

The software was tested using members of the research team as participants in order to gain data for analysis to uncover the possibilities to do a larger scale study at a later stage and to uncover what type of results can be found using such a framework. Before the trial, all watches that were used were booted with a clean installation of the latest version of MERGEN. This means that all settings had been set to the same and was not possible to be changed by the participants in the trial. The sample rate was set to be based on time rather than location as the emphasis in this trial was on the biometric data rather than the position. In reality, this means that several data points might be collected in the same position if the person is static. This is preferable over the option of having a location-based collection which would collect a data point every X meter as the result of heart rate changing at the same location might be of interest when searching for external factors affecting travellers and would be lost if data were only to be collected every X meter. It was also possible to set the quality of the data collected. The software was set to collect fine location data, when possible, and coarse data when fine location data was not available. The sample rate of the collection was set to one sample per minute as this would interfere as little as possible with the battery life of the device, based on benchmarks made by the developers of the framework.

## 3.2 Data analysis

In the previous subsection, the method for data collection was described. This subsection aims to explain what types of data analysis methods were explored and how the results were extracted and compared. The goal of the analysis was to uncover areas where the heart rate deviates from its regular patterns. The hypothesis was that it would be possible to uncover how travellers are affected by certain areas through the combined utilisation of several data analysis methods. The methods used for the analysis is described in this section and provides a background for the results which will be presented in the following section.

The data was refined using Microsoft Excel, and all cluster-based analysis methods were explored using MATLAB R2019a with the “Statistics and Machine Learning Toolbox” as well as the “Mapping Toolbox”. The regression-based analysis was performed using SPSS 26. The analysis was conducted in a linear fashion, where each step either led to the refinement of the input data for the next step or the results for comparison between the methods. Each step will be described as in-depth in the following paragraphs.

### 3.2.1 Data refinement

The data was limited to a border containing location points within and around the Stockholm area in Sweden. As to not affect the comparison between different persons’ patterns, the heart rate was normalised between 0 and 1. No further data refinement was performed before the analysis.

### 3.2.2 K-means clustering

The data analysis was first explored using K-means clustering to find groupings of areas where the biometric data, in this case, heart rate, was deviating from the regular pattern. After considering the finding from the literature review, it was decided that the Average Silhouette Evaluation (ASE) method was going to be used to determine  $K$ , rather than the common elbow method, as the elbow method can be considered more ambiguous than the ASE method since it sometimes can be troublesome to choose between two integers, leading to a suboptimal  $K$ .

However, the K-means clustering does not put any weight on any of the dimensions and therefore compares all data points with the same weight per dimension. As the heart rate has a smaller range of possible values than the position, the clustering tended to organise the data according to the spatial positions rather than the locations. For this reason, the ASE method was used solely on the heart rate dimension to find the appropriate number of clusters.

Once  $K$  was defined, the data was clustered using all dimensions and K-means clustering.

### 3.2.3 DBSCAN

The K-means clustering organised the data into levels of heart rate deviation and unveiled where deviations of heart rate could be observed. However, to find more significant areas where travellers were affected, a clustering of locations also had to be done. In order to do so, a method for clustering the data regarding positioning was used. In this case, DBSCAN was decided to be suitable as it would also filter out noise points which were of no interest when comparing larger areas.

The clusters from the K-means clustering was used to create the density-based clusters. Each cluster was utilised as input in DBSCAN and the output where several areas per cluster that could be singled out and used for future comparison. However, the clusters did not implicate any information about the correlations between different factors. Therefore, linear regression was used to analyse the data and uncover if there was any correlation between the condition of the person and the specific areas.

### 3.2.4 Creating the Personal Biometric Model

The exploration of the data shows that there seems to be some spatial effect on the biometric data of the participant. In order to further analyse and understand how individual components affect the collected biometric data, a model had to be created and tested. To create the model, a comparison of the collected variables and prominent factors that could affect heart rate was performed and led to the creation of an array of variables that should be included in the model. The variables included spatial, temporal, biometric and movement data which could be divided into the linear and categorical features that can be seen in Table 2.

Table 2 - The variables for the model

<i>Input</i>	<i>Derived Variables</i>
<i>PPG Sensor (Heart Rate Sensor)</i>	Heart rate (linear)
	Speed (linear)
<i>GPS/GLONASS</i>	Points of Interest (categorical)
	Elevation (linear)
<i>Time</i>	Type of day (categorical)
	Time of day (categorical)
	Movement in X axis (linear)
<i>Accelerometer Data</i>	Movement in Y axis (linear)
	Movement in Z axis (linear)

The data for the heart rate was normalised between 0-1 using MATLAB and the three accelerometer axes X, Y and Z were divided as AX, AY and AZ. The rest of the independent variables required some form of data handling which is described in 3.2.4.1-3.2.4.4. The creation of the model can be found in 3.2.4.5.

### 3.2.4.1 Speed

In order to get the speed, the distance between the current and the last collected data point had to be found and divided by the time between them. To calculate the distance between two positional points that are collected with decimal degree coordinates, one has to account for the radius of earth in order to find a distance that resembles the ground truth as closely as possible. However, the radius of earth is not consistent and fluctuates between 6353 km and 6384 km depending on where on the surface of earth the data has been collected, because of the geoidal bulge. This does not account for the above mean sea level (AMSL), and for the sake of this trial, the mean of 6371 km was determined as the mean radius of the earth. In order to determine the speed by each position, the speed was measured from the last point to the current point, as to reflect the speed that had affected the heart rate that was collected at the current point. The linear data of speed was divided to have six different coefficients in order to better understand the activity of the person. The person was deemed to be stationary between 0-1 km/h, the coefficient for this would be 0, so this coefficient would not be part of the model. The person was deemed to be walking ( $SV_1$ ) if the recorded speed was 1-5 km/h, brisk walking ( $SV_2$ ) if the speed was 5-10 km/h, running ( $SV_3$ ) if the speed was 10-15 km/h, running really fast (athletic) ( $SV_4$ ) if the speed was 10-20 km/h, and anything faster than 20 km/h was deemed to be caused by the use of a vehicle ( $SV_5$ ).

To calculate the distance between two positioning points in the data set, the haversine formula was used. The haversine formula calculates the distance between two decimal degree coordinates and accounts for the radius of the earth. The calculation can be seen in Equation 2 to 4 below. Assuming that:

$$\varphi_1 = \text{Latitude point 1} \quad \varphi_2 = \text{Latitude point 2}$$

$$\lambda_1 = \text{Longitude point 1} \quad \lambda_2 = \text{Longitude point 2}$$

$$R = \text{the radius of Tellus; 6371 km}$$

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos\varphi_1 * \cos\varphi_2 * \sin^2\left(\frac{\Delta\lambda}{2}\right) \quad (\text{eq. 2})$$

$$c = 2 * \text{atan2}\left(\sqrt{a}, \sqrt{(1-a)}\right) \quad (\text{eq. 3})$$

$$d = R * c \quad (\text{eq. 4})$$

### 3.2.4.2 Points of Interest

For the sake of this trial, three types of point of interest were determined. Those were “office” ( $POI_1$ ), “home” ( $POI_2$ ) and “other” based on geographical boundaries around the visit locations. The boundaries for each location was decided by creating a rectangle around the office and the home.

Table 3 - The boundaries for the POIs

Boundaries	$POI_1$ Office	$POI_2$ Home
Upper Latitude Limit	59.344838	59.259986
Lower Latitude Limit	59.353255	59.263847
Upper Longitude Limit	18.076021	18.079773
Lower Longitude Limit	18.054048	18.087132

### 3.2.4.3 Elevation

To get the elevation, each position had to be compared to a dataset containing elevation information (Becker et al., 2009). However, the resolution of said datasets varies, meaning that the resolution of the data in the elevation dataset might not overlap perfectly with the data from the trial dataset. In order to account for this, an algorithm was written to find the closest location which had elevation data and utilise that data for the trial data point. This is a form of interpolation and which just copied the evaluation value of the closest data point. The resolution of the elevation dataset was 30 m.

### 3.2.4.4 Time Categories

The type of day was categorised into weekdays ( $D$ ) and not weekdays. For the time of day, three categories were chosen; “morning” ( $T_1$ ) which was from 6:00 to 11:59, “afternoon” ( $T_2$ ) which was 12:00-17:59 and “evening/night” which was 18:00-05:59. These categories were selected based on different time-space constraints, and subsequently level of stress, that an individual may have on a given typical day (Liu et al., 2018; Susilo and Liu, 2017).

### 3.2.4.5 Personal Biometric Model

The variables from the derived data were put together to create the Personal Biometric Model (PB-model) for effects of sensor readings on collected heart rate. The model was based on the variables described above. Time of day ( $T$ ), type of day ( $D$ ), Elevation and location ( $POI$ ) were used to control temporal and spatial factors. The rest of the data relating to the movement were also put in the model and a constant ( $C$ ) was added as everyone should have a base heart rate.

Thus, based on the variables listed above, the Personal Biometric Model (PB-model) employed is as below:

$$NormHR = C + (\alpha_1SV_1 + \alpha_2SV_2 + \alpha_3SV_3 + \alpha_4SV_4 + \alpha_5SV_5)Speed + \beta_1AX + \beta_2AY + \beta_3AZ + \gamma Elevation + \zeta_1POI_1 + \zeta_2POI_2 + \eta D + \theta_1T_1 + \theta_2T_2 \quad (eq. 5)$$

Based on the format of the model, it was possible to run it using multivariate linear regression to find the coefficients for each variable in the model, which will be presented in the results.

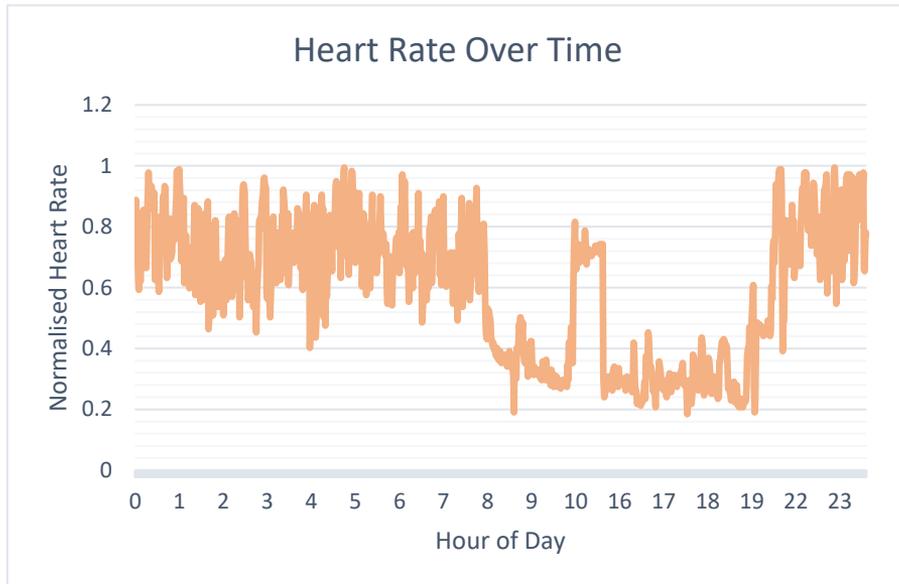
## 4 Results

This section aims to visualise and present the results from the data collection and analysis described in the previous section. The results will be described in three sections depicting each step of the analysis and the results from each.

### 4.1 Data collected

Data was collected from 5 participants over the course of approximately three months. However, the consistency of the data was varying between the participants as some of the participants forgot about the collection and others did modifications to the hardware which interrupted the collection of data. In the end, the one participant’s dataset which was most consistent and contained the most data points were chosen to be the dataset for the analysis part of this trial. Over the course of approximately one month, 12495 data points were collected from said participant, a 43-year-old male. During this month, the participant travelled both domestically and

abroad, which gave many data points to analyse. The participant's heart rate fluctuation of a regular day can be seen in Figure 7.



*Figure 7 - A diagram over the heart rate of the participant over the course of a day*

As we can see in Figure 7, the heart rate is fluctuating throughout the day. However, the patterns during evening/night seem to differ a lot from the rest of the day. There can be many reasons as to why, but one reason could be that the participant is not wearing the device correctly (or has even taken it off) and static noise from the environment has been collected instead. The software was supposed to control for when the device was not worn correctly and would then not collect any data, but the possibility of the device collecting noise instead of data can still not be ruled out without further testing. However, in the time spans defined as morning and afternoon, the data seems to follow a more regular pattern, with the occasional peak. Hopefully this is an indication that correct data has been collected during that time, and for the sake of this experimental work, the assumption will be that the amount of data will be large enough for the analysis methods to disregard any errors.

#### 4.1.1 Data refinement

The data that had been collected was saved to several CSV-files and had to be compiled to one. The software that collected the data also saved data that was not relevant for this trial, such as gyroscope data. The CSV-files were combined to one large CSV-file including all data. The redundant data was deleted, and the heart rate was normalised as to not affect the results when comparing the findings with those

of other participants in the future. The date and time stamps of the data had to be individual in order to work with the methods of this trial. The data was filtered to only include data points from the area around Stockholm, where the majority of the data points were collected, which resulted in a data set of 7650 points when the boundaries were set to those in Table 3.

*Table 4 - Decimal Degree Coordinate Boundaries*

*Boundaries*

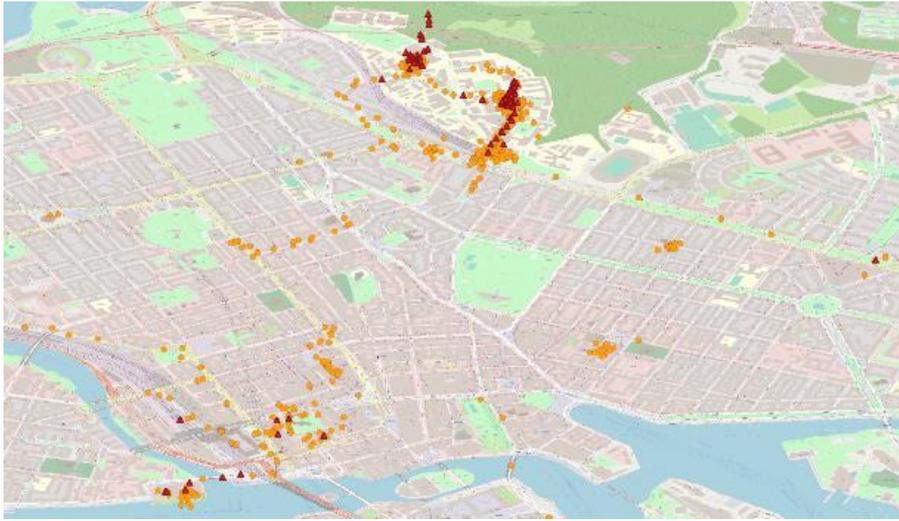
<i>Upper Latitude Limit</i>	60
<i>Lower Latitude Limit</i>	56
<i>Upper Longitude Limit</i>	20
<i>Lower Longitude Limit</i>	16

## 4.2 Data analysis

The analysis was performed as described in the methodology, and in this section, the result from each step will be presented.

### 4.2.1 K-means clustering

The first step of the data analysis was to find the integer to use as  $K$  for K-means clustering. This was performed with the use of the Average Silhouette Evaluation method on the whole data set, and  $K$  was determined to be 2. With  $K$  found, the dataset was clustered using the K-means clustering algorithm. The two clusters were then visualised on a map over Stockholm as can be seen in Figure 8.



*Figure 8 - A visualisation of the clusters. Orange circles indicate regular heart rate; Red triangles indicate peak heart rate*

## 4.2.2 DBSCAN

After clustering and visualising the data with K-means clustering, the two data clusters were divided into several clusters using DBSCAN, which also removed any noise and outliers from the dataset in order to find areas where the data collection was dense. In the cluster with regular heart rate, there were 25 spatial clusters and one layer of outliers, and in the cluster, with peak heart rate there were five spatial clusters. These spatial clusters were then compared to see where there were overlaps between the clusters and where it was not. Three of the deviation clusters overlapped with the regular pattern clusters, uncovering that in those areas the heart rate fluctuated. In 22 of the regular heart rate clusters, there was no overlap, and in two of the deviation clusters there was no overlap.

## 4.2.3 Personal Biometric Model

The data was also put through the PB-model in order to uncover how much each factor contributed to the heart rate readings. It showed that there had been some peaks of speed in the data which did not make sense (speeds at several thousand meters per second) and most were interpolated with the surrounding values. However, there were 2 points of data which had speed peaks that would not get interpolated, so those two points were removed from the dataset as not to disturb the outcome of the multivariate regression using the PB-model. The results are shown in Table 5.

Table 5 - The results of the multivariate regression on the PB-model

<i>Regression Statistics</i>			
<i>Multiple R</i>	0.719		
<i>R Square</i>	0.517		
<i>Adjusted R Square</i>	0.516		
<i>Standard Error</i>	0.149		
<i>Observations</i>	7648		
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
<i>Intercept</i>	1.303	0.193	6.745
<i>Walking</i>	0.004	0.009	0.445
<i>Brisk Walking</i>	0.010	0.008	1.344
<i>Running</i>	0.003	0.007	0.421
<i>Marathon</i>	-0.001	0.004	-0.123
<i>Faster</i>	-0.002	0.001	-2.587
<i>AX</i>	-0.002	0.000	-6.012
<i>AY</i>	0.012	0.001	23.611
<i>AZ</i>	-0.008	0.000	-28.093
<i>Elevation</i>	-0.839	0.192	-4.379
<i>POI<sub>1</sub>(Office)</i>	0.076	0.005	16.127
<i>POI<sub>2</sub>(Home)</i>	0.295	0.010	28.943
<i>D(Weekdays)</i>	-0.047	0.004	-11.717
<i>T<sub>1</sub>(Morning)</i>	0.020	0.005	3.974
<i>T<sub>2</sub>(Afternoon)</i>	-0.026	0.005	-4.845

## 5 Discussion

The previous section has presented the results found using the methodology described in this paper. This section aims to explain what the results implicate, both in general and how the results matter for traffic and urban planners. This section also aims to discuss how the quality of data and the methods for data collection might have affected the results and why that matters for future trials of similar sorts like the one presented in this paper.

### 5.1 The clustering in general

The clustering of data showed that it is possible to find locations and areas where the heart rate is consistent over time as well as locations and areas where the heart rate deviates. The K-means clustering visualisations indicate that an increased amount of collected data points in an area give more opportunity for peak heart rates to be collected in the same area. This is further confirmed with the density-based clustering as we can see that areas with more significant amounts of collected data experience both deviations and regular heart rate. Further data collections should be performed in these areas with added dimensions relating to the level of stress or cognitive load the participant is experiencing. Many factors can affect how the heart rate increases and decreases in a single area, and these factors need to be found and accounted for.

However, the results of the PB-model better describe how the participant was affected and which factors had a play in it. In the following section 5.2, the results of the PB-model will be discussed, and in section 5.3, there will be a discussion on the implications of these findings on urban planning.

### 5.2 The results of the PB-model

The results of the PB-model showed how the different variables affected the normalised heart rate. What is interesting to see is that the speed of the person is rather insignificant in the sense that its t-value is within the range for rejecting the hypothesis that they might be affecting the heart rate. For this dataset, that range would be -1.960 to 1.960 for 95% confidence. However, what has to be taken into consideration is that these states are only compared to the stationary state and the data consisted of many more data points where the participant was stationary than mobile. It is also interesting to see that the higher speeds are affecting the heart rate negatively, meaning that the faster the participant has travelled, the lower the heart rate has become. By looking at the pattern of the speed coefficients, it is indicative that there seems to be a threshold where this happens. A reason for why it might be like that is that once a person has gotten into a vehicle, in theory the person should no longer be affected by the speed itself. However, a person is still susceptible to the

changes in speed while in a vehicle and for future iterations of the model, both the difference in speed and acceleration should be included.

The accelerometer data indicates the gravitational pull on the participant's device, which could be used as an indicator for the intensity of the activities being conducted. For this data collection, the sample rate was set to one sample per minute or 0.017 Hz, but with a higher sample rate, it would not only be possible to better understand how much a person is moving but also what type of activity is being performed if gyroscopic data was added as well. Both the software MERGEN and the PB-model should be updated to include this as it would give a better indication of how much the activity of the person is influencing the heart rate. With the data collected, however, it is possible to see that the readings are rather significant by looking at their t-values, which might mean that there is more information to be found here.

The significance of the elevation also seems to be rather high. For this iteration of the software, no elevation data was collected with the rest of the data; it was rather gathered by comparing the spatial data to registers of elevation at specific locations. This means that only elevation changes outdoors were accounted for and no elevation changes within a building could be collected. However, the significance of this variable unsurprisingly shows that there is more information to be gathered and explored related to elevation. A variable such as elevation change since last collected data point would give a better explanation of how the elevation is affecting the person. The absolute values being used now only shows that higher altitudes, in general, lower the heart rate of a person, which could be because of thinner air or other reasons.

The Points of Interests (POI) were used to show how a location can affect a person, and both POIs of the model are compared to a third category called "other" which is neither home nor office. It is interesting to see that the person's heart rate is positively affected by both being located at home and at the office compared to other locations. Naively, this could be interpreted as if the person has an easier time being away from home and office since the heart rate would be lower, but that could be since the "other" category contains every other location possible within the realm of the data collection. So, POIs within the "other" category could include locations which could be able to lower the heart rate, such as spas or similar POIs. This means that a broader range of POIs could be useful to find how the heart rate is affected by them. This is particularly interesting for urban planners as they can choose their POIs as the locations where they will change or have changed the built environment to see whether or not the impact on the heart rate has increased or lowered or is the same. More about this will be discussed in 5.3.

The weekdays and times of day are all significant variables that affect the heart rate of a person. The reason as to why the heart rate would be lower during weekdays than weekends might be because when the person is working at their office, they are

less physically active than in their weekends where they might have activities which require more physical activities, such as going for walks or performing different forms of sport.

The interpretations could be further investigated by performing interviews with the participant to see whether or not they might be correct, something which could be implemented in future trials.

### 5.3 Implications for urban planning

With the results found in this trial, the framework can be found useful for traffic and urban planners. By collecting the biometric data at more locations over a longer time it will be possible to find areas that have had participants experience both regular heart rate patterns and deviations. By analysing these areas and finding the correlation between the activities performed in the area and the heart rate it will be possible to rule out which factors, such as walking or running, that are affecting the person. When finding several areas with the same characteristics it will then become possible to compare the features of those areas until correlations between said features and the heart rate patterns can be found. Building up a data library of such features can then help in planning future areas where the planners want no, or encouraging, effects from the built environment on the inhabitants or travellers. With the dawn of smart cities, where every citizen can be connected to smart services, the utilisation of such frameworks as the one trialled in this paper will provide additional information on how cities can be planned to be less intrusive on the wellbeing of the inhabitants as the technology becomes available to a broader audience than before.

However, the unique selling point which needs to be addressed is the fact that this data is collected directly from the participant without any psychological interpretations handling the data before it is being collected. This is beneficial of several reasons, mainly two which are; 1) the removal of the human factor by collecting data directly from the source leads to fewer mistakes which could have created inadequate or faulty data, and 2) the removal of the need to interpreting questions from surveys also eliminates interpretation errors which could have caused data that that could have been either biased or completely faulty. By measuring biometric data and referring to methods for psychophysiological extraction of mental states and conditions (Hosseini and Khalilzadeh, 2010; Jerčić et al., 2018; Lee et al., 2004; Setz et al., 2010; Sharma and Gedeon, 2013; Wijsman et al., 2013; Zhai and Barreto, 2006), one could reveal more reliable results given that the hardware and software used is of the best quality possible (von Watzdorf and Michahelles, 2010).

## 5.4 Risks

All the data analysis in this paper has been performed based on the assumption that the data that has been collected, has been collected correctly and that all sensors have been calibrated accordingly during the manufacturing process. That being said, there can, of course, be deviations in the data that is caused by misuse or faulty equipment, and there is no safe way to tell if this has happened. However, all deviations in the data have, for the sake of the data analysis in this trial, been treated as deviations in the sources, meaning that heart rate deviations in the collected data have been treated as ground truth deviations in the participant's heart rate.

The elevation data used in the data analysis had a much lower resolution than the data collected in this trial. By finding the closest value and using that, there might be many data points that share the same elevation data, when in reality the participant has walked uphill and downhill between two data points, which cannot be discovered using the current dataset. That being said, there is, of course, the possibility to redo the analysis with higher resolution data, but there will not be any possibility to collect ground truth elevation data with the use of the chosen equipment for this trial.

By using time-dependent data collection, it has been made sure that no heart rate changes are missed in a static location, however, depending on the speed of the traveller, increases and decreases in heart rate while travelling quickly might be missed. For future trials, a fusion of location-based and time-based collection can be used, where the software collects a sample every  $X$  meter and  $Y$  seconds, whichever comes first.

## 6 Conclusion & Future Work

The work described in this paper shows the possibility to collect and analyse the position and biometric data collected with the software chosen for this trial.

- Machine learning methods, such as K-means clustering and DBSCAN, seems like a viable solution for analysis of the collected data to find locations of interest.

The PB-model can further be used to find correlations between factors in each area at specific times, to see if the physical changes are affecting the heart rate or if other types of factors related to the environment are affecting it.

- The tool is proven useful in the current state for urban planners, as they can be informed of how previous planning decisions are affecting the wellbeing of the inhabitants of an area by looking at their biometric data.

Future directions of this work should include extensive data collection in order to have a more significant data set to analyse. However, the trial presented in this paper has given valuable information regarding the current possibilities to collect and analyse information regarding the effects of external factors on the participant. In order to create a more accessible, and sustainable, society, models such as the PB-model can give indications of how the urban landscape should be formed in order to include all possible target groups.

## 7 Acknowledgements

This work was supported by NOESIS (NOvel Decision Support tool for Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services, <http://noesis-project.eu/>), Project number 769980, EU H2020 funded project.

## 8 References

Allström, A., Kristoffersson, I., Susilo, Y., 2017. Smartphone based travel diary collection: experiences from a field trial in Stockholm. *Transp. Res. Procedia* 26, 32–38.

American Heart Association, 2015. Electrocardiogram (ECG or EKG) | American Heart Association [WWW Document]. URL <https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/electrocardiogram-ecg-or-ekg> (accessed 5.20.19).

Apple Inc, 2019a. Apple Watch Series 3 - Technical Specifications [WWW Document]. URL [https://support.apple.com/kb/sp766?locale=en\\_US](https://support.apple.com/kb/sp766?locale=en_US) (accessed 5.22.19).

Apple Inc, 2019b. watchOS 5 - Apple [WWW Document]. URL <https://www.apple.com/watchos/watchos-5/> (accessed 8.5.19).

Apple Inc, 2019c. App Store - Apple [WWW Document]. URL <https://www.apple.com/ios/app-store/> (accessed 8.5.19).

ASICS Digital Inc, 2019. Runkeeper - Track your runs, walks and more with your iPhone or Android phone [WWW Document]. URL <https://runkeeper.com/> (accessed 8.6.19).

Axhausen, K.W., Zimmermann, A., Schönfelder, S., Rindsfuser, G., Haupt, T., 2002. Observing the rhythms of daily life: A six-week travel diary. *Transportation (Amst)*. 29, 95–124.

Becker, J.J., Sandwell, W.H.F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., von Rosenberg, J., Wallace, G., Weatherall, P., 2009. Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30\_PLUS. *Mar. Geod.* 32, 355–371.

Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D.C., Pensabene, V., Actis, P., 2018. Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLoS One* 13, e0192691.

Bohte, W., Maat, K., 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transp. Res. Part C Emerg. Technol.* 17, 285–297.

Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp. Res. Part C Emerg. Technol.* 68, 285–299.

Empatica, 2019. Real-time physiological signals | E4 EDA/GSR sensor [WWW Document]. URL <https://www.empatica.com/en-int/research/e4/> (accessed 5.22.19).

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A Density-Based Algorithm for

Discovering Clusters in Large Spatial Databases with Noise.

Ferrer, S., Ruiz, T., 2014. Travel Behavior Characterization Using Raw Accelerometer Data Collected from Smartphones. *Procedia - Soc. Behav. Sci.* 160, 140–149.

Fitbit, 2019a. Fitbit Charge 3 | Advanced Fitness Tracker [WWW Document]. URL <https://www.fitbit.com/charge3> (accessed 5.22.19).

Fitbit, 2019b. Technology [WWW Document]. URL <https://www.fitbit.com/technology> (accessed 8.5.19).

Gadziński, J., 2018. Perspectives of the use of smartphones in travel behaviour studies: Findings from a literature review and a pilot study. *Transp. Res. Part C Emerg. Technol.* 88, 74–86.

Galton, F., 1886. Regression Towards Mediocrity in Hereditary Stature. *J. Anthropol. Inst. Gt. Britain Irel.* 15, 246–263.

Garmin Ltd., 2019. vivoactive 3 Element | Wearables | Products | Garmin | Singapore | Home [WWW Document]. URL <https://www.garmin.com.sg/products/wearables/vivoactive-3-element-cerise/> (accessed 8.5.19).

Gonzalez, P.A., Weinstein, J.S., Barbeau, S.J., Labrador, M.A., Winters, P.L., Georgi, N.L., Perez, R., 2010. Automating mode detection for travel behaviour analysis by using global positioning systems-enabled mobile phones and neural networks. *IET Intell. Transp. Syst.* 4, 37.

Google LLC, 2019a. Wear OS overview | Android Developers [WWW Document]. URL <https://developer.android.com/training/wearables/> (accessed 5.22.19).

Google LLC, 2019b. Google Play [WWW Document]. URL <https://play.google.com/store> (accessed 8.5.19).

Gorny, A.W., Liew, S.J., Tan, C.S., Müller-Riemenschneider, F., 2017. Fitbit Charge HR Wireless Heart Rate Monitor: Validation Study Conducted Under Free-Living Conditions. *JMIR mHealth uHealth* 5, e157.

Healthy Doc, 2018. How Does PPG Technology Works? - SoulFitBlog [WWW Document]. URL <https://soulfit.io/blog/how-does-ppg-technology-works/> (accessed 5.20.19).

Ho Yu, C., 1977. Exploratory Data Analysis. *Methods* 2, 131–160.

Hosseini, S.A., Khalilzadeh, M.A., 2010. Emotional Stress Recognition System Using EEG and Psychophysiological Signals: Using New Labelling Process of EEG Signals in Emotional Stress State. In: 2010 International Conference on Biomedical Engineering and Computer Science. IEEE, pp. 1–6.

Huawei Technologies Co., 2019. HUAWEI WATCH 2, NFC, Google Assistant,

- 4G SIM, sport watch |HUAWEI Global [WWW Document]. URL <https://consumer.huawei.com/en/wearables/watch2/> (accessed 5.22.19).
- Jerčić, P., Sennersten, C., Lindley, C., 2018. Modeling cognitive load and physiological arousal through pupil diameter and heart rate. *Multimed. Tools Appl.* 1–15.
- Jovanov, E., Raskovic, D., Price, J., Chapman, J., Moore, A., Krishnamurthy, A., 2001. Patient Monitoring Using Personal Area Networks of Wireless Intelligent Sensors. *Biomed. Sci. Instrum.* 37, 373–378.
- Kassambara, A., 2018. Determining The Optimal Number Of Clusters: 3 Must Know Methods - Datanovia [WWW Document]. Datanovia. URL <https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/> (accessed 8.6.19).
- Kaufman, L., Rousseeuw, P.J., 1990. Finding groups in data : an introduction to cluster analysis. Wiley-Interscience.
- Lai, T.L., Robbins, H., Wei, C.Z., Hannan, E.J., 1979. Strong Consistency of Least Squares Estimates in Multiple Regression II\*, *JOURNAL OF MULTIVARIATE ANALYSIS*.
- Lee, M., Yang, G., Lee, H.-K., Bang, S., 2004. Development stress monitoring system based on personal digital assistant (PDA). In: *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 2364–2367.
- Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., Cheng, T., 2016. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS J. Photogramm. Remote Sens.* 115, 119–133.
- Li, X., Ge, M., Dai, X., Ren, X., Fritsche, M., Wickert, J., Schuh, H., 2015. Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo. *J. Geod.* 89, 607–635.
- Liu, C., Susilo, Y.O., Dharmowijoyo, D.B.E., 2018. Investigating intra-household interactions between individuals' time and space constraints. *J. Transp. Geogr.* 73, 108–119.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: *Lecam, L., Neyman, J. (Eds.), Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.* University of California Press, Berkeley, pp. 281–297.
- MEILI - travel diary [WWW Document], 2017. URL <https://appadvice.com/app/meili-travel-diary/1308940609> (accessed 3.15.18).
- Murakami, E., Wagner, D.P., 1999. Can using global positioning system (GPS) improve trip reporting? *Transp. Res. Part C Emerg. Technol.* 7, 149–165.

Numenta, 2014. Geospatial Tracking Learning the Patterns in Movement and Detecting Anomalies Geospatial Tracking Executive Summary Advances in Anomaly Detection [WWW Document]. URL [https://numenta.com/assets/pdf/whitepapers/Geospatial Tracking White Paper.pdf](https://numenta.com/assets/pdf/whitepapers/Geospatial%20Tracking%20White%20Paper.pdf) (accessed 1.6.18).

Palmberg, R., Susilo, Y., Gidofalvi, G., 2019a. Uncovering Effects of Spatial and Transportation Elements on Travellers Using Biometric Data. In: Toivonen, T., Karst, G., Willberg, E. (Eds.), TOWARDS HUMAN SCALE CITIES - OPEN AND HAPPY. Department of Geosciences and Geography, University of Helsinki, Helsinki, p. 125.

Palmberg, R., Susilo, Y., Gidofalvi, G., 2019b. Developing and trialling an implicit interaction platform to monitor and aiding dementia travellers. In: Mobile Apps and Sensors in Surveys (MASS) Workshop. Mannheim.

Prelipcean, A.C., Gidófalvi, G., Susilo, Y.O., 2014. Mobility Collector. *J. Locat. Based Serv.* 8, 229–255.

Prelipcean, A.C., Gidófalvi, G., Susilo, Y.O., 2017. Transportation mode detection – an in-depth review of applicability and reliability. *Transp. Rev.* 37, 442–464.

Prelipcean, Adrian C., Gidófalvi, G., Susilo, Y.O., 2018a. MEILI: A travel diary collection, annotation and automation system. *Comput. Environ. Urban Syst.*

Prelipcean, Adrian Corneliu, Susilo, Y.O., Gidofalvi, G., 2018. Future directions of research for automatic travel diary collection.

Prelipcean, Adrian C., Susilo, Y.O., Gidófalvi, G., 2018b. Collecting travel diaries: Current state of the art, best practices, and future research directions. *Transp. Res. Procedia* 32, 155–166.

Prelipcean, A.C., Yamamoto, T., 2018. Workshop Synthesis: New developments in travel diary collection systems based on smartphones and GPS receivers. *Transp. Res. Procedia* 32, 119–125.

Rashidi, T.H., Abbasi, A., Maghrebi, M., Hasan, S., Waller, T.S., 2017. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transp. Res. Part C Emerg. Technol.* 75, 197–211.

Reddy, S., Burke, J., Estrin, D., Hansen, M., Srivastava, M., 2008. Determining transportation mode on mobile phones. In: 2008 12th IEEE International Symposium on Wearable Computers. IEEE, pp. 25–28.

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M., 2010. Using mobile phones to determine transportation modes. *ACM Trans. Sens. Networks* 6, 1–27.

Rencher, A.C., Christensen, W.F., 2012. Multivariate Regression. In: *Methods of Multivariate Analysis*. John Wiley & Sons, Ltd, pp. 339–383.

- Samsung, 2019a. Gear S3 classic [WWW Document]. URL <https://www.samsung.com/us/mobile/wearables/smartwatches/sm-r770nzsaxar-sm-r770nzsaxar/> (accessed 5.22.19).
- Samsung, 2019b. Galaxy Watch (46mm) Silver (4G LTE) [WWW Document]. URL <https://www.samsung.com/us/mobile/wearables/smartwatches/galaxy-watch--46mm--silver--lte--sm-r805uzsaxar/> (accessed 5.22.19).
- Setz, C., Arrich, B., Schumm, J., La Marca, R., Troster, G., Ehlert, U., 2010. Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Trans. Inf. Technol. Biomed.* 14, 410–417.
- Sharma, N., Gedeon, T., 2013. Hybrid Genetic Algorithms for Stress Recognition in Reading. In: *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics. EvoBIO 2013. Lecture Notes in Computer Science, Vol 7833.* Springer, Berlin, Heidelberg, Berlin, pp. 117–128.
- Silbert, S., 2019. All the Things You Can Track With Wearables [WWW Document]. Lifewire. URL <https://www.lifewire.com/what-wearables-can-track-4121040> (accessed 8.5.19).
- Simjanoska, M., Gjoreski, M., Gams, M., Madevska Bogdanova, A., 2018. Non-Invasive Blood Pressure Estimation from ECG Using Machine Learning Techniques. *Sensors (Basel)*. 18.
- Stopher, P.R., Daigler, V., Griffith, S., 2018. Smartphone app versus GPS Logger: A comparative study. *Transp. Res. Procedia* 32, 135–145.
- Susilo, Y.O., Liu, C., 2017. Exploring patterns of time-use allocation and immobility behaviours in the Bandung Metropolitan Area, Indonesia. *Urban Mobilities Glob. South* 131–153.
- TRavelVU [WWW Document], 2016. URL <https://www.trivector.se/it-system/programvaror/travelvu/> (accessed 3.15.18).
- von Watzdorf, S., Michahelles, F., 2010. Accuracy of positioning data on smartphones. In: *Proceedings of the 3rd International Workshop on Location and the Web - LocWeb '10.* ACM Press, New York, New York, USA, pp. 1–4.
- Wallen, M.P., Gomersall, S.R., Keating, S.E., Wisløff, U., Coombes, J.S., 2016. Accuracy of Heart Rate Watches: Implications for Weight Management. *PLoS One* 11, e0154420.
- Wang, Z., He, S.Y., Leung, Y., 2018. Applying mobile phone data to travel behaviour research: A literature review. *Travel Behav. Soc.* 11, 141–155.
- Wijsman, J., Grundlehner, B., Liu, H., Penders, J., Hermens, H., 2013. Wearable Physiological Sensors Reflect Mental Stress State in Office-Like Situations. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction.* IEEE, pp. 600–605.

Zhai, J., Barreto, A., 2006. Stress Recognition Using Non-invasive Technology. In: FLAIRS Conference. pp. 395–401.

Zhang, T., 2004. Solving Large Scale Linear Problems Using Stochastic Gradient Descent Algorithms. In: Proceedings of the Twenty-First International Conference on Machine Learning. SCM, pp. 116–123.