WEATHER EFFECTS ON HUMAN MOBILITY: AN ANALYSIS USING MULTI-CHANNEL SEQUENCE ANALYSIS

Vanessa S. Brum-Bastos[1], Jed A. Long[1], Urška Demšar[1]
[1] School of Geography & Sustainable Development, University of St Andrews,
Irvine Building, North Street, St Andrews, UK
Email: {vdsbb*, jed.long, urska.demsar}@st-andrews.ac.uk

*Corresponding author

Abstract

Widespread availability of geospatial data on movement and context presents opportunities for applying new methods to investigate the interactions between humans and weather conditions. Understanding the influence of weather on human behaviour is of interest for diverse applications, such as urban planning and traffic engineering. The effect of weather on movement behaviour can be explored through Context-Aware Movement Analysis (CAMA), which integrates movement geometry with its context. More specifically, we use multi-channel sequence analysis (MCSA) to represent a person's movement as a multi-dimensional sequence of states, describing either the type of movement or the state of the environment throughout time. Similar movement patterns can then be identified by comparing and aligning mobility sequences. In this paper we apply CAMA and MCSA to explore weather effects on human movement patterns. Data from a GPS tracking study in a Scottish town of Dunfermline are linked to weather data and converted into multi-channel sequences which are clustered into groups of similar behaviours under specific weather typologies. Our findings show that the CAMA + MCSA method can successfully identify the response of commuters to variations in environmental conditions. We also discuss our findings on how travel modes and time spent at different places are affected by meteorological conditions, mainly wind, but also rainfall, daylight duration, temperature, comfort and relative humidity.

Keywords: context-aware movement analysis, context-aware similarity, human mobility, human movement, multi-channel sequence analysis, context.

1. Introduction

The spread of geolocated smartphones and the decreasing price of GPS devices have contributed towards the production of large amounts of data on human movement of unprecedented spatio-temporal quality (Meekan et al., 2017). New human mobility studies attempt to link such movement data with contextual information (such as points of interest) to gather insights into, for example, commuting behaviour (Beecham, Wood, & Bowerman, 2014; Gong, Chen, Bialostozky, & Lawson, 2012), tourist behaviour (Meijles, de Bakker, Groote, & Barske, 2014; Versichele, Neutens, Delafontaine, & Van de Weghe, 2012), or retail choice decisions and human activities (Siła-Nowicka et al., 2016). However, integrating high resolution GPS trajectories and dynamic spatio-temporal contextual information remains an underexplored approach for studying the effects of weather on human movement, despite its relevance for urban planning (Givoni, 1974; Ng, 2012), traffic engineering (Dunne & Ghosh, 2013), retail planning (Thakuriah, Sila-Nowicka, & Paule, 2016), tourism (de Freitas, 2003), health (Tucker & Gilliland, 2007), psychology (Nerlich & Jaspal, 2014) and epidemiology (Horowitz, 2002).

Specific weather conditions often trigger changes in human behaviour, for example, higher temperatures increase aggressiveness (Anderson, 2001; Carlsmith & Anderson, 1979) and lower temperatures contribute to irritability and combativeness (Schneider, Lesko, & Garret, 1980; Worfolk, 1997). Different components of weather have different magnitudes of importance, for example, air temperature, direct solar radiation and wind speed have a more significant influence on human behaviour than humidity (de Montigny, Ling, & Zacharias, 2012). However, it is challenging to understand how weather influences human behaviour because the responses are partially a result of individual preferences (de Freitas, 2015). Some individuals are more responsive to the thermal component of weather, i.e. the combined effects of air temperature, humidity and solar radiation, while some are more receptive to physical components like rain, and others are more greatly affected by the aesthetic components, such as cloud coverage and sunshine. Yet, most individuals do respond to the combination of all three of these components (de Freitas, 1990).

Traditionally, these interactions have been explored through questionnaires and multidimensional scaling methods within the field of human biometeorology (Cabanac, 1971; de Freitas, 1990; Manu, Shukla, Rawal, Thomas, & de Dear, 2016). With the increased availability of tracking and environmental data we however propose that the effect of weather on movement behaviour can be explored through Context-Aware Movement Analysis (CAMA), which integrates movement geometry with its context, i.e. with the surrounding biological and environmental conditions that might be affecting movement (Andrienko, Andrienko, & Heurich, 2011; Demsar et al., 2015; Dodge et al., 2013). More specifically we use multi-channel sequence analysis (MCSA) to represent a person's movement as a sequence of states, describing either the type of movement or the state of the environment throughout time. Similar movement patterns can then be identified (termed context aware similarity analysis) by comparing and aligning mobility sequences.

Similarity analysis is one of the most common tasks in movement analytics and consists of using distance measures and grouping methods to split trajectories (Demšar et al. 2015) into groups of elements more similar amongst them than to other groups (Jain et al. 1999), which followed by clustering allows the identification of spatio-temporal movement patterns that might be linked to behaviour (Dodge, Weibel, Ahearn, Buchin, & Miller, 2016). Similarity is often established based on geometry or physical attributes; geometrical similarity solely relies on measures of spatial and temporal distances, and physical similarity relies on movement attributes such as speed, turning angle, acceleration and direction (Demsar et al., 2015). Context-aware similarity is based on multiple attributes (Andrienko et al., 2011; Buchin, Dodge, & Speckmann, 2014; Demsar et al., 2015; Sharif & Alesheikh, 2017b) describing the conditions within which the movement took place.

Context-awareness is a recent trend (Sharif & Alesheikh, 2017a), as a result there are few context-aware methods for assessing similarity between trajectories. Sharif & Alesheikh (2017b) generalized the dynamic time warping (DTW) to develop a context-based dynamic time warping (CDTW) method, which matches trajectories with contextual similarity even if they are not concurrent. This method is highly dependent on arbitrary weights for the contextual variables, restricted to numeric context and disregards changes of context between two points in time. i.e., same contexts are considered similar even when they are not concurrent. De Groeve et al. (2016) uses single channel sequence alignments and Hamming Distance to understand the temporal variation of habitat use by roe deer; the similarity is measured by the cost to transform a sequence of habitat use into another. This method is able to handle only one contextual variable at time, therefore it is not able to handle the interactive effect of multiple contextual variables on movement. Buchin et al. (2014) modified existing similarity measures to make them context-aware, more specifically they defined the distance between two points as the sum of their contextual and spatial distances. The transition costs between contexts are defined by the user and the method is restricted to contextual data in the form of polygonal divisions.

In this paper we propose to use multi-channel sequence analysis (MCSA) to perform context-aware similarity analysis (CASA) and cluster trajectories into groups of similar behaviour. MCSA is a new analysis tool for movement data where contextual information can now be readily combined with detailed tracking datasets. The main advantage of this approach is that it also is possible to consider as many channels (contextual variables) as desired at once. It is common in movement research to simultaneously consider multiple environmental variables, which makes MCSA particularly relevant for studying human mobility, traffic, transportation and wildlife ecology; areas in which movement behaviour may be contextualised by other dynamic environmental variables such as air temperature, vegetation indices, humidity, wind speed, air pollution and snow coverage. Single channel analysis has been used before to explore spatio-temporal patterns on the activity of visitors in Akko's Old city –

Israel (Shoval & Isaacson, 2007) and to analyse sequential habitat use by roe deer in North-East Italy (De Groeve et al., 2016). Shoval & Isaacson (2007) focused on sequences of locations, i.e. the movement itself, while De Groeve et al (2016) emphasized sequences of habitat use classes, i.e. the context surrounding movement. Horanont *et al.* (2013) looked at GPS traces from mobile phone users, coarse scale movement data, hourly temperature, rainfall and wind speed to explore the independent effects of each variable on people's activity patterns. We innovate by applying MCSA, for the very first time, to perform CAMA of fine scale human movement data to simultaneously consider movement and context by looking at the combined and single effects of six meteorological variables.

Despite the novelty of MCSA in movement research, sequence analysis has been consistently used in medical and social sciences, particularly within bioinformatics and life courses research (Idury & Waterman, 1995) Abbott 1995; Abbott & Tsay 2000). In bioinformatics, a sequence represents the DNA molecule as a string of characters (which stand for specific nucleotides), between a precise start and end point; the comparison of similarities and differences between those strings allows the identification of nucleotide sequences related to genetic diseases and traits. We propose that the same principle can be applied to movement trajectories for identifying groups of people with similar movement patterns, i.e., clusters of similar behaviour (Billari, 2001). Further, we propose to not only represent the trajectories with one sequence only, but to use Multi-channel sequence analysis (MCSA), which allows for comparison of sequences consisting of several dimensions (channels) (Gauthier *et al.*, 2010). For this, we link data from a GPS tracking study to weather data and convert the information into multi-channel sequences in a first fully data-driven attempt to explore weather effects on human movement patterns.

The rest of the paper is structured as follows: first we describe the GPS tracking data and weather datasets used in our analysis. Next, we explain how the meteorological data sources were combined and integrated with the GPS tracking data and finally converted into sequences. Next, multi-channel sequence analysis is applied to identify changes in group movement patterns related to weather. We conclude with considerations on our findings, the potential of the methodology and ideas for future research.

2. Methodology

To study the influence of weather on human mobility behaviour we used a five-step process (Figure 1). In Step 1, we integrate trajectories with contextual data by using trajectory annotation to link GPS points to weather variables, which resulted in contextualized trajectories. In Step 2, we transform those trajectories into multi-channel sequences by creating alphabets with codes for each weather variable, travel mode and places. In Step 3, we use optimal matching distances (Abbott & Tsay, 2000) to calculate a dissimilarity matrix describing the degree of difference between each pair of multi-channel sequences in our dataset. In Step 4, we use Ward's clustering (Murtagh & Legendre, 2011) algorithm to partition the sequences into similarity based groups, which represent groups of people showing similar movement behaviour under particular weather conditions. In Step 5, we perform statistical test to validate and understand differences between groups.

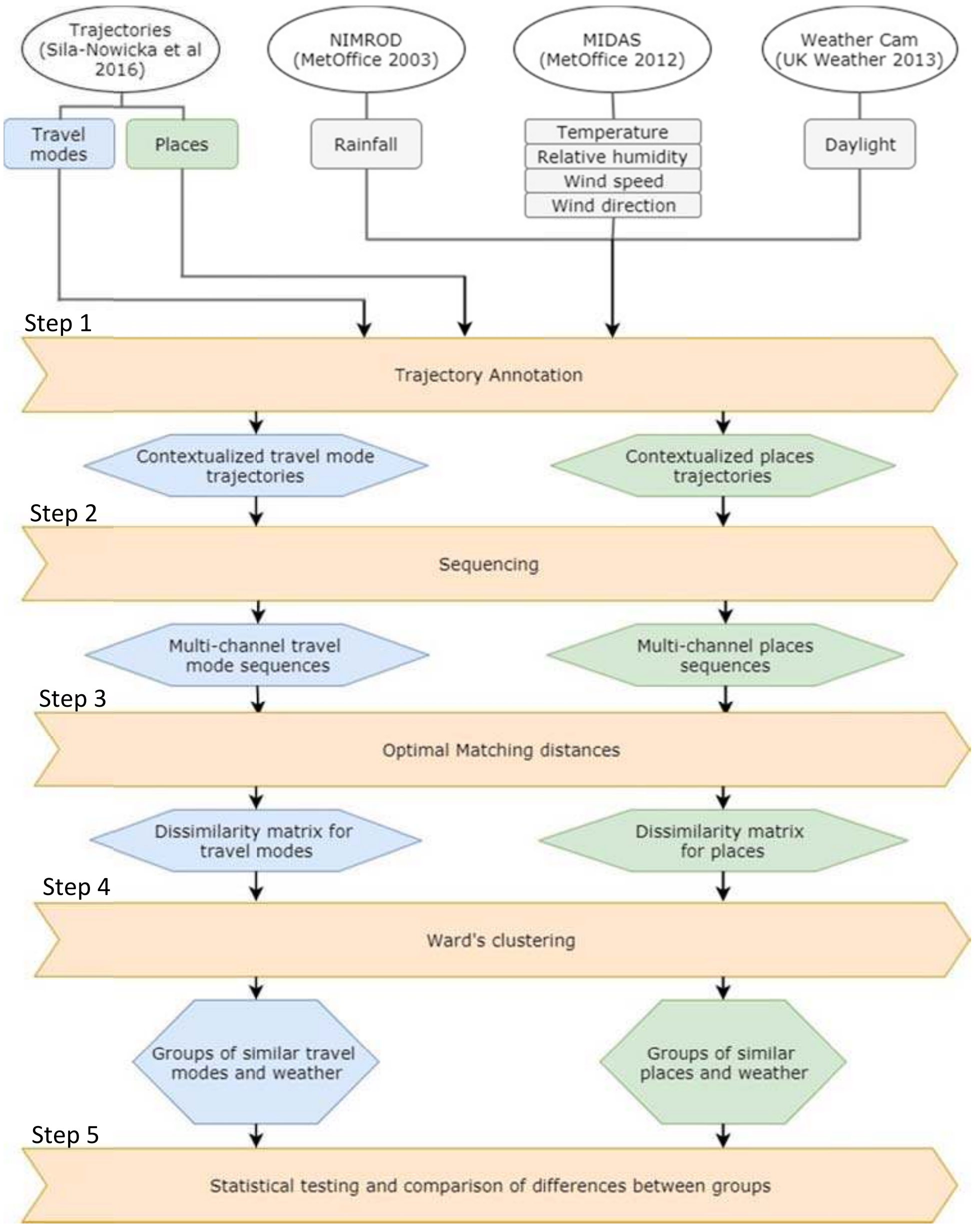

Figure 1 – The overview of our framework for identification of groups of similar movement behaviour under specific weather conditions. The framework has two analyses running in parallel: analysis of places and analysis of travel modes. Blue shapes marks travel mode, green shapes marks places, white ellipses represent dataset's sources, rectangles represent variables, beige arrows represent processing steps and hexagons derived results in each step.

Trajectory annotation and sequencing were performed using PostgreSQL 9.4 database manager, VANJU library and its dependencies under Python 2.7, for more details refer to Brum-Bastos, Long, & Demšar (2016). The MCSA, including optimal matching distances, Ward's clustering and statistical tests, was performed using TraMineR 1.8-9 and cluster 1.14.4 libraries under R 3.4.1, for more details on the equations used by these libraries please refer to Gabadinho, Ritschard, Studer, & Müller (2009) and Maechler, Rousseeuw, Struyf, Hubert, & Hornik, (2018) respectively.

2.1. Movement data

We analysed a human movement dataset where GPS devices were carried by volunteers from the Kingdom of Fife – UK (Figure 2a) (Siła-Nowicka et al., 2016). The data were collected between the 28th of September 2013 and the 10th of January 2014 as part of the GEOCROWD project (Siła-Nowicka et al., 2016), in which 6000 individuals were randomly selected by postcode address from the voting registry (focusing on the three major towns in this region) and invited via letter to participate in the study. In total, 206 individuals accepted the invitation and provided useable data whereby they were tracked for two consecutive weeks within the study time spam. GPS devices recorded participant positions every five seconds, representing a very high-resolution trajectory of participant locations. The GPS trackers were coupled with accelerometers, which turned off the GPS when the individual was not moving Oshan *et al.* 2014). The aim of the GEOCROWD project was to develop new movement analytics methods that would allow researchers to find out as much as possible from the actual GPS data while participants were asked to do as little as possible (i.e. the only task was to carry a GPS device and mail it back after two weeks). Therefore, very little auxiliary data were collected and beyond gender and age of the participants, which were sourced from the electoral register together with the address of each participant, no other demographic or ground truth data were collected. For more details on data collection refer to Oshan *et al.* (2014).

In this paper we re-analyse the GEOCROWD data from the town (called Dunfermline; Figure 2a) with highest number of participants (n=91), of which 23 were female, 41 were male, and 27 did not declare their gender. Looking at the ages of our participants: 10 were between 21 and 34 years old, 46 were between 35 and 60 years old, 8 were between 61 and 65 years old, and 27 did not declare their age. As stated above, apart from their home address, gender, and age; no further information about participants or their activities were available for our secondary data analysis.

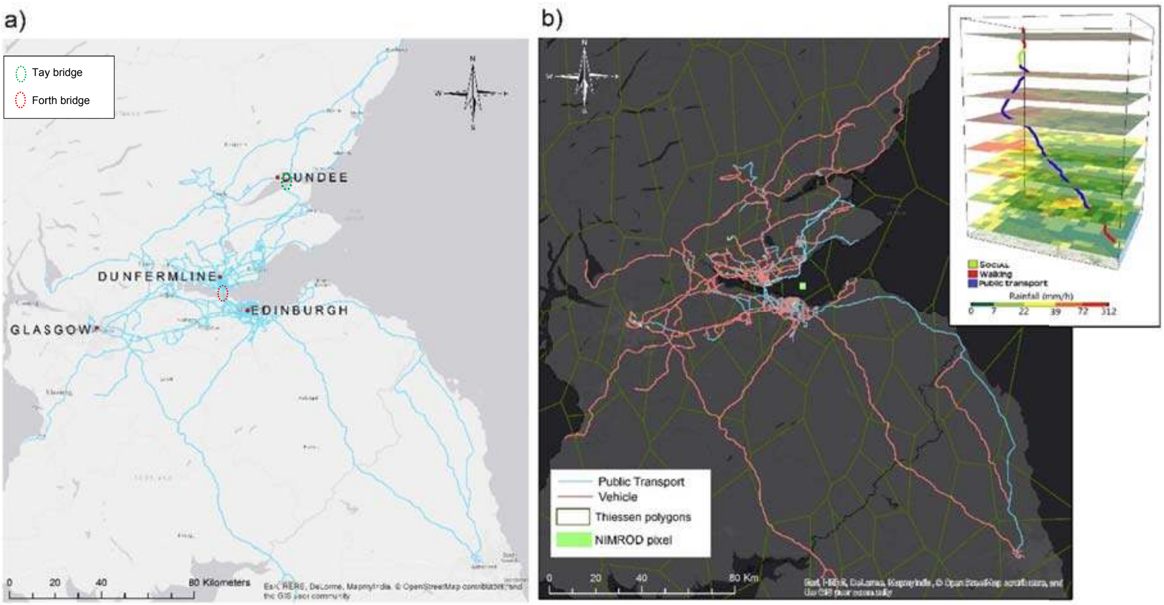

Figure 2 – a) East coast of Scotland where the GPS data were collected, and trajectories represented by light blue lines. The green and the red ellipses represent the locations of the Tay and Forth road bridges respectively. b) Sample of two movement travel modes overlaid by the Thiessen polygons used to interpolate MIDAS (Met Office Integrated Data Archive System) data and one pixel of a NIMROD (Met Office's nowcasting system) product for comparison. The frame in the right upper corner illustrates a trajectory sample classified into movement modes and displayed in a space-time cube with rainfall data for a one-hour period.

The participant trajectories were classified into movement classes (Walk, Train, Bus and Vehicle, Traffic Stop, Bus Stop, Train Stop, Fig. 2b) and stop classes (Home, Work, Shopping, Unidentified Stop) (Siła-Nowicka et al., 2016). The classification achieved 85% accuracy, which was assessed by comparing a 200 m range from the recorded home addresses with the home location found by the classification algorithm (for more details on data segmentation and classification refer to Sila-Nowicka *et al.* 2016).

2.2. Contextual data and context integration

We linked meteorological data from ground stations and orbital satellites to movement data through linear dynamic trajectory annotation (DTA-L) (Brum-Bastos et al., 2016), a method that estimates the contextual variable at the time when the GPS point was collected by interpolating the values immediately before and after the point chronologically. The DTA-L method accounts for the rate-of-change between contextual layers, producing more realistic values for interpolated meteorological data, and it also deals with the difference between temporal resolutions of the datasets (Brum-Bastos et al. 2017). We collated multiple sources of contextual data on weather (Table 1).

Table 1– Contextual datasets with respective sources and specifications.

Source	***Variables***	***Data type***	***Geometry***	***Temporal resolution***	***Spatial resolution***
Weather Cam (UK Weather, 2013)	Daylight	Categoric	Point	24 h	---
NIMROD (MetOffice, 2003)	Rainfall	Numeric	Raster	5 min	1-5 km
MIDAS (Met Office, 2012)	Temperature Relative humidity Wind speed Wind direction	Numeric	Point	1 h	---

We associated MIDAS data with trajectory points using Thiessen Polygons around each meteorological station (n = 109, Fig 2b). From the MIDAS meteorological variables we also derived the apparent temperature (AT), which considers the combined effects of temperature, humidity and wind (Steadman, 1994).

$$AT = Ta + 0.33 * e - 0.70 * Ws - 4.00 \tag{1}$$

Here Ta is the air temperature in °C; e is the water vapour pressure in hPa calculated from the relative humidity and temperature; and Ws is the wind speed in m/s.

The Weather Cam data was used to calculate dusk, sunset, sunrise and dawn times (for a central location in the study area) as at this latitude daylight length varies by approximately 4.5 hours from September to January. Daylight categories were annotated to trajectories according to the following rules: Morning Twilight (MT) for fixes recorded in the period between dawn and sunrise, Day Light (DL) for fixes recorded between sunrise and sunset, Evening Twilight (ET) for fixes recorded between sunset and dusk, Night (NI) for fixes recorded between dusk and dawn.

2.3. Trajectory sequencing

Sequence analysis requires a finite alphabet, in which each letter originally represents genomic nucleotides (Idury & Waterman, 1995). In single channel sequence analysis, a sequence is a one dimensional ordered list of characters from one alphabet, representing successive states (Abbott &

Tsay, 2000). However, most phenomena are multidimensional and require multiple alphabets. This means that each dimension gets its own bespoke alphabet and instead of having the data object represented as one sequence, the object now has as many different sequences as there are dimensions, which are called channels (therefore the name Multi-Channel Sequence Analysis). The alignment, i.e. similarity, then needs to be calculated across all channels along the time axis (Gauthier et al., 2010). This multi-channel approach is therefore a shift from looking at individual units towards analysing context, connections and events (Abbott, 1995).

We created several bespoke alphabets, one for movement mode (e.g., walking and driving) and one for each weather variable in our data. For this, we had to translate the GPS track of each participant into a multi-channel sequence consisting of time units, to which the characters were assigned (figure 3). Weather conditions were categorized to create weather-based alphabets (Table 2). Rainfall was classified based on the UK Met Office scale for rainfall intensity, Wind Speed according to an adaptation of the Beaufort scale (Royal Meteorological Society, 2017), wind direction according to the cardinal and collateral points, apparent temperature according to the VDI (2008) thermal perception scale, humidity and temperature according to the 1991-2000 seasonal climate normals for Dunfermline from Jenkins *et al.* (2009). Climate normals are a three-decade average of weather variable commonly used to characterize local climates (Ayoade, 1986).

The multi-channel sequences were then generated for each volunteer and day (illustrated in Figure 3) by taking the modal weather condition (for each variable described in Table 2) and movement mode for each 1-minute interval for each participant. To each time unit we assigned descriptors for the weather variables and the respective movement mode, which are linked to the descriptor for the following time unit building multiple chronologically arranged strips. These sequences can be analysed alongside strips of contextual variables to understand not only the responses to specific variables, but also to different combinations of those variables and the identification of patterns relative for specific age groups, gender or other profiling information. The number of channels in a MCSA is defined by the number of variables under consideration, in our case eight variables therefore, eight channels by definition. The use of modal attributes for each 60 second segment (as the data were collected at a 5 second frequency) filtered out possible noise from the raw data and represents an appropriate scale of analysis for studying human movement.

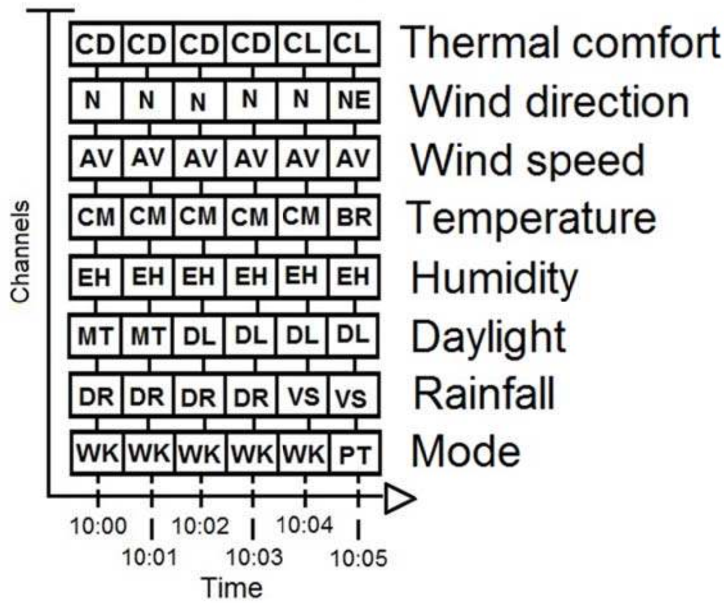

Figure 3 - A multi-channel sequence for a participant over a five-minute period, each channel relates to one of the meteorological variables and movement modes for that minute of the day.

We calculated the entropy index (EI) for the movement mode channel for all sequences of at each minute (Billari, 2001). The EI is a measure of the complexity induced by the distribution of states in a group of sequences (Gabadinho et al., 2009),which in our case can be used to observe the diversity of places and travel modes across the week and hours of the day. In our analysis, an EI closer to one indicates an even distribution of a contextual variable across movement modes (alphabet states), while an EI closer to zero indicates a high level of association with one mode. We also looked at the average time expenditure at home, socialising, shopping, walk, public transport and vehicle by gender and on each day of the week. The average time expenditure was calculated by first computing the amount of time spent in each movement mode and dividing it by the total GPS active time for each participant, keeping in mind that each state in our sequences corresponded to one minute. Following this, we calculated the mean for the gender of participants (male, female).

Table 2 – Alphabets for meteorological variables used as contextual data with respective ranges and description. Letters in each alphabet are defined based on standard meteorological classifications (see text for more details).

Thermal perception (°C)			*Rainfall (mm/h)*			*Wind coming from direction (°)*		
Letter	**Description**	**Range**	**Letter**	**Description**	**Range**	**Letter**	**Description**	**Range**
VC	Very Cold	<= -39	**DR**	Dry	0	**N**	North	> 337.5 - 22.5
CD	Cold	>-39 - -26	**VS**	Very Slight	>0 - 0.5	**NE**	North East	>22.5 -67.5
CL	Cool	>-26 - -13	**SL**	Slight	>0.5 – 1	**E**	East	> 67.5 -112.5
SC	Slightly Cool	>-13 - 0	**LM**	Low Moderate	>1 – 2	**SE**	South East	>112.5 - 157.5
CF	Comfortable	>0 - 20	**MO**	Moderate	>2 – 4	**S**	South	>157.5 – 202.5
SW	Slightly Warm	>20 - 26	**HV**	Heavy	>4– 10	**SW**	South West	>202.5 – 247.5
W	Warm	>26 – 32	**VH**	Very Heavy	>10 - 50	**W**	West	> 247.5 – 292.5
H	Hot	>32 – 38	**VI**	Violent	> 50	**NW**	North West	>292.5 – 337.5
VH	Very Hot	>38						
Humidity (%)			*Temperature (°C)*			*Wind Speed (m/s)*		
Letter	**Description**	**Range**	**Letter**	**Description**	**Range**	**Letter**	**Description**	**Range**
EH	Extremely High	>90	**EL**	Extremely Low	<=5	**CM**	Calm	<= 3
AA	Above Average	>85 -90	**AN**	Average minimum	> 5-7	**BR**	Breeze	>3 - 14
AV	Average	>80 -85	**AV**	Average Average	>7 -10	**GA**	Gale	>14 - 24
BA	Below Average	>75 -80	**AX**	Average Maximum	>10 - 13	**ST**	Storm	>24
LW	Low	>70 -75	**EH**	Extremely High	>13			
EL	Extremely Low	<70						

2.4. Context-Aware Similarity Analysis (CASA)

2.4.1. Multi-channel Sequence Analysis (MCSA)

We divided our analysis into two streams, by separately analysing travel modes (walk, public transport and vehicle) and places (home, social places and shopping), since the choice of travel mode and of staying in a place are not necessarily affected in the same way by weather (Derrick Sewell, Kates, & Philips, 1968). When the destination is obligatory, such as work, people are more likely to change their travel mode, for example, driving to work instead of walking under heavy rain; however, if the destination is linked to leisure, such as shopping, people might simply postpone the task instead of changing the travel mode to get there (Connolly, 2008; Zivin, 2014). We further split the analysis into weekdays and weekends to reflect different movement motivations (for example, travel to work during workdays is usually obligatory regardless of weather conditions while people have more voluntary choices about their mobility during weekends).

Sequence analysis requires cost matrices, which were computed separately for travel modes and places and for weekends and weekdays. We used the optimal matching (OM) distance to compute similarity between sequences as this method has shown potential for identifying groups with matching movement behaviour (De Groeve et al., 2016). The distance between two sequences is assessed by quantifying their differences based on a matrix with the costs for substituting, deleting or inserting letters to transform one sequence into the other. The substitution costs are given by symmetrical matrices that represent the costs of transitioning between each pair of states in the alphabet (Gabadinho et al., 2009). In our case, the costs for transitions between the states of travel modes, places, wind speed and wind direction were computed using transition rates calculated from the sequences for computing the cost matrices, as shown in Equation 2.

$$F(i,j) = 1 - P(i,j) - P(j,i) \quad (2)$$

Here $F(i,j)$ is the substitution cost, $P(i,j)$ is the transition rate from state i to j.

The costs for transitions between the states of thermal comfort, temperature, humidity, daylight and rainfall were defined by ordering the classes of each variable (alphabets) by their intensity and calculating the cost to replace one class by another with Equation 3.

$$F(i_n, j_{n+1}) = \frac{|n-(n+1)|}{z-1} \quad (3)$$

Here $F(i_n, j_{n+1})$ is the cost between the classes i and j with intensity order n and $n+1$, and z is the number of classes for that variable (size of the alphabet). The cost for replacing null values by any other class (insertion) was zero and likewise to substitute any other class by null (deletion), because for our study they are related to periods for which we had no information on the participant's movement. This procedure resulted in ten cost matrices, one for each weather variable, two for travel modes and two for places (weekdays and weekend).

The cost matrices are then used to calculate the optimal match (OM) score, for example, given an alphabet A with size Z, pick sequences I and J based on alphabet A. The sequences are aligned in time and the OM cost is calculated by summing up the costs of substitutions ($C_{S_iS_j}$), deletions and insertions (d) needed to modify the sub sequences of J, so that it turns into I. The OM is the less costly and is computed using Equation 4, in which each line defines a possible OM score for two sub sequences, depending on which of the procedures, insertion, deletion or substitution, is cheaper (Gauthier et al., 2010).

$$F(i,j) = min \begin{cases} F(i-1, j-1) + C_{S_iS_j} \\ F(i-1, j) + \mathrm{d} \\ F(i, j-1) + \mathrm{d} \end{cases} \tag{4}$$

Here $F(i-1, j-1)$ represents the OM score of a subsequence containing the 1 to $i-1$ characters of sequence I against a subsequence containing 1 to $j-1$ in sequence J (Gabadinho et al., 2009; Gauthier et al., 2010). The OM cost is computed for each channel between all multi-channel sequences and the cost between two multi-channel sequences is the summed costs between their channels. We calculated the OM distances simultaneously considering three channels for wind: movement mode, wind speed and wind direction; and two channels for the remaining weather conditions, where the places or the travel modes were always the first channel and the variables were considered in turns as the second channel. A k by k dissimilarity matrix, where k is the number of sequences, represents the level of alignment between each two multi-channel sequences, i.e., a similarity measure between two moving people.

2.4.2. Cluster analysis & typology

The dissimilarity matrix can be used to find whether people were showing similar movement behaviour under certain weather conditions. For this we apply a clustering algorithm to the dissimilarity matrix for each weather variable for both travel modes and places. We used Ward's clustering, a hierarchical bottom-up algorithm that computes dissimilarities between two groups as the increase in the error sum of squares after merging those groups. The algorithm starts with each sequence as their own group and successively merges them into clusters based on the minimum increase in the error sum of squares, until it becomes a single cluster (Murtagh & Legendre, 2011). For selecting the optimal number of clusters, we used the Calinski-Harabaz Index (CHI) (Calinski & Harabasz, 1974) that considers the within and between groups dispersion as shown in Equation 5.

$$CHI = \frac{trace\ (B)}{trace\ (W)} \tag{5}$$

Here W and B are the within and between group dispersion matrices, the trace of W is the sum of the within cluster variance and the trace of B is the sum of the between cluster variances; a higher CHI indicates a better data partition (Ahmed, 2012), because it shows that the within group distances are lower and the between groups distances are higher. We varied the number of clusters from the number of sequences (i.e. the maximum possible number of clusters, if every sequence is allocated to its own cluster) to one and used the configuration with highest CHI, except where the maximum CHI resulted in individuals' clusters, to assign the multi-channel sequences into their final clusters. The combination of values of weather and movement modes in each cluster then defined a type of the group. Note that the types are not consistent between variables, i.e., we found different clusters for each weather variable, thus the typology is specific for each variable.

We then looked at the distribution of the proportion of time spent in different travel modes and places for the weather conditions associated with each cluster. We expected this would give insights into the different behavioural patterns in individuals related to weather (i.e., an overall picture of the effects of the weather conditions within each cluster on movement modes). We tested the significance of the differences using Kruskal-Wallis and Levene's tests and we assumed that a statistically significant difference between medians or variances of each cluster was enough evidence to support the existence of different behavioural groups.

We further used discrepancy analysis to verify if and how behavioural groups were related to age and gender. This method evaluates the strength of the association between the groups of sequences and a categorical covariate (Studer, Ritschard, Gabadinho, & Müller, 2011) by calculating the share of discrepancy according to Equation 6 and looking at its p-value.

$$SD = \frac{SS_B}{SS_T} \quad (6)$$

Here SD is the share of discrepancy, SS_B is the sum of square distances within the age or gender groups, and SS_T is the total sum of square distances between all sequences (Batagelj, 1988).

3. Results

3.1. Trajectory sequencing

The different movement modes for each participant for each day of the week are shown in Figure 4, most sequences start between six and eight in the morning and have a minimum of 58% and a maximum of 98% of missing data, i.e., minutes for which the GPS tracker was off and movement modes are unknown (white gaps in Figure 4). The entropy index (EI) (Figure 5) provides some insight into participants daily movement behaviour. The EI increases between 4:00 am and 7:00 am on weekdays, but only rises between 8:00 am and 10:00 am on weekends, indicating higher diversity of movement modes earlier on weekdays. Sunday has the highest EI and similarly to Saturday, it drops and rise between 3:00 pm and 6:00 pm.

The average time spent (AVTS) walking did not change substantially across weekdays and between genders (Figure 6). The AVTS at home varied throughout the week, being the highest on Sunday and lowest on Wednesday for both genders (dashed orange lines on Figure 6). The low values on Wednesday might be related to the higher average time spend socializing in comparison to other days of the week (dashed red lines on Figure 6). Moreover, women seem to spend more time socializing and to concentrate social activities on Tuesdays, Wednesdays and Saturdays; while men socialise very little on Tuesdays and keep a steady, but lower than women, average from Wednesday to Monday.

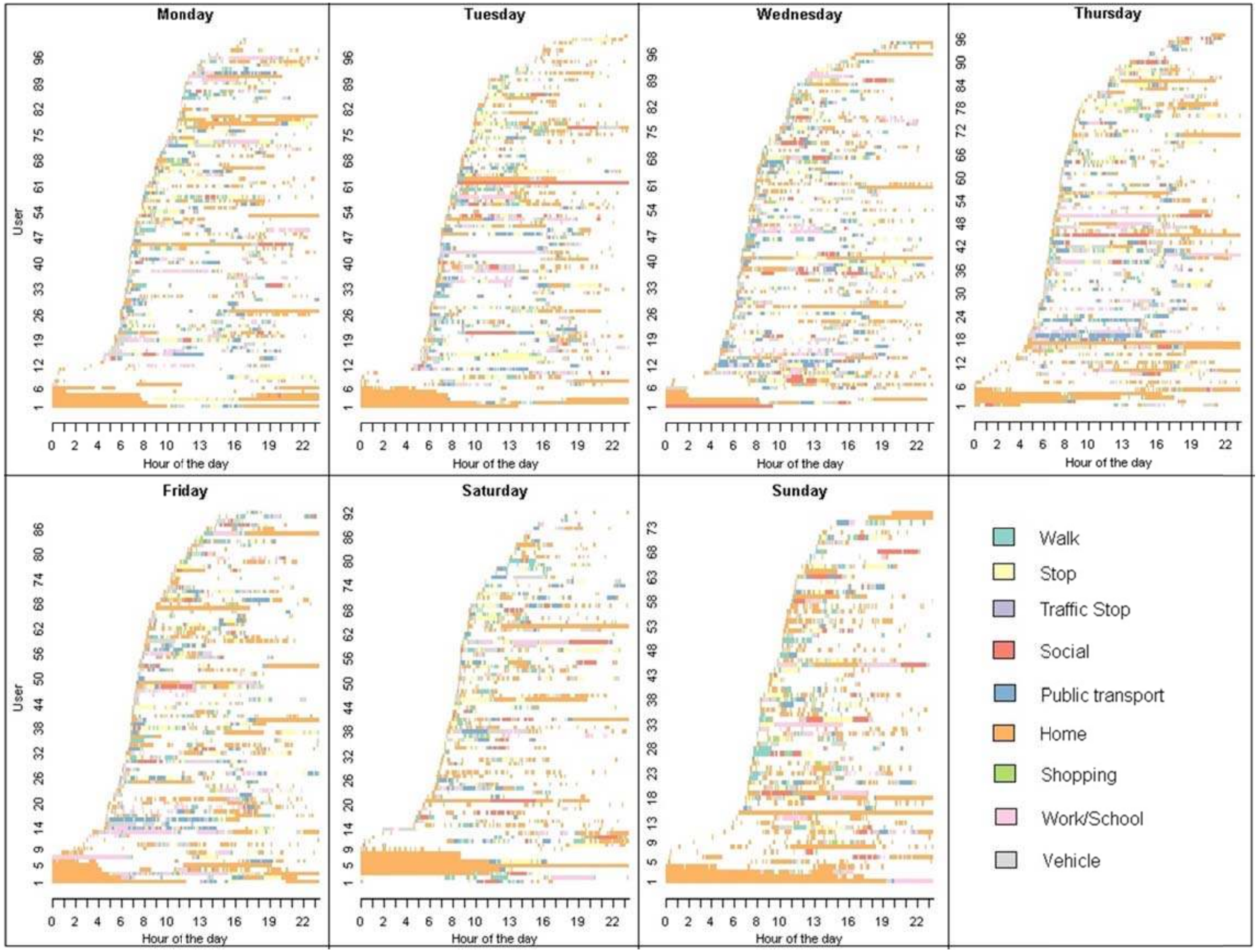

Figure 4 - Sequences of movement modes for each day of the week by participant (vertical axis) and by hour of the day (horizontal axis), missing data are reported in white. For each day, the sequences are ordered by length, but orderings are different for each day (that is, the sequence number on the y chart does not always indicate the same individual).

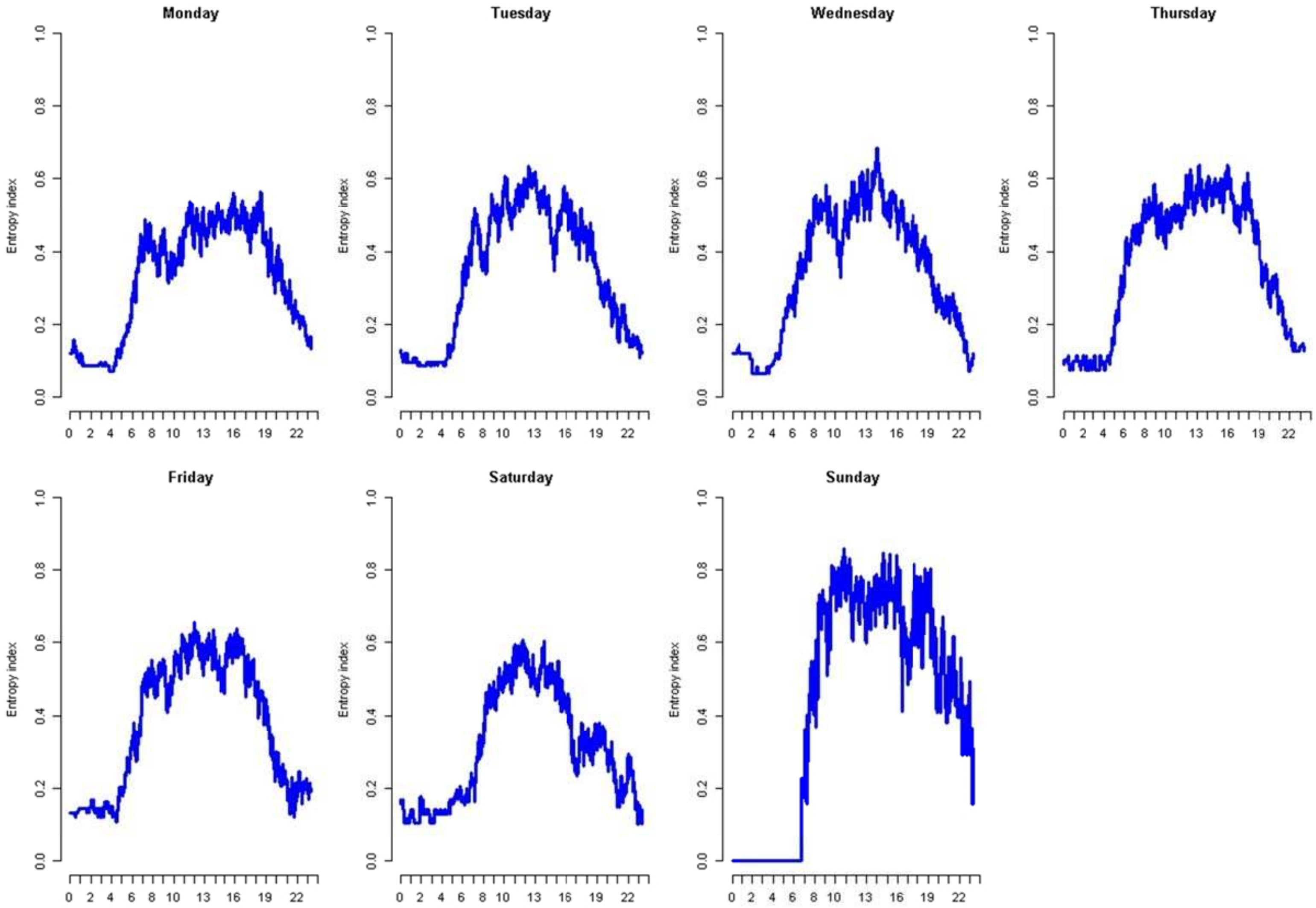

Figure 5 - Entropy index (EI) by day of the week and by hour of the day (horizontal axis), an EI close to zero indicates that most people are in the same movement mode at that time (lower diversity), while an EI close to one indicates similar proportion of people on the different movement modes (higher diversity).

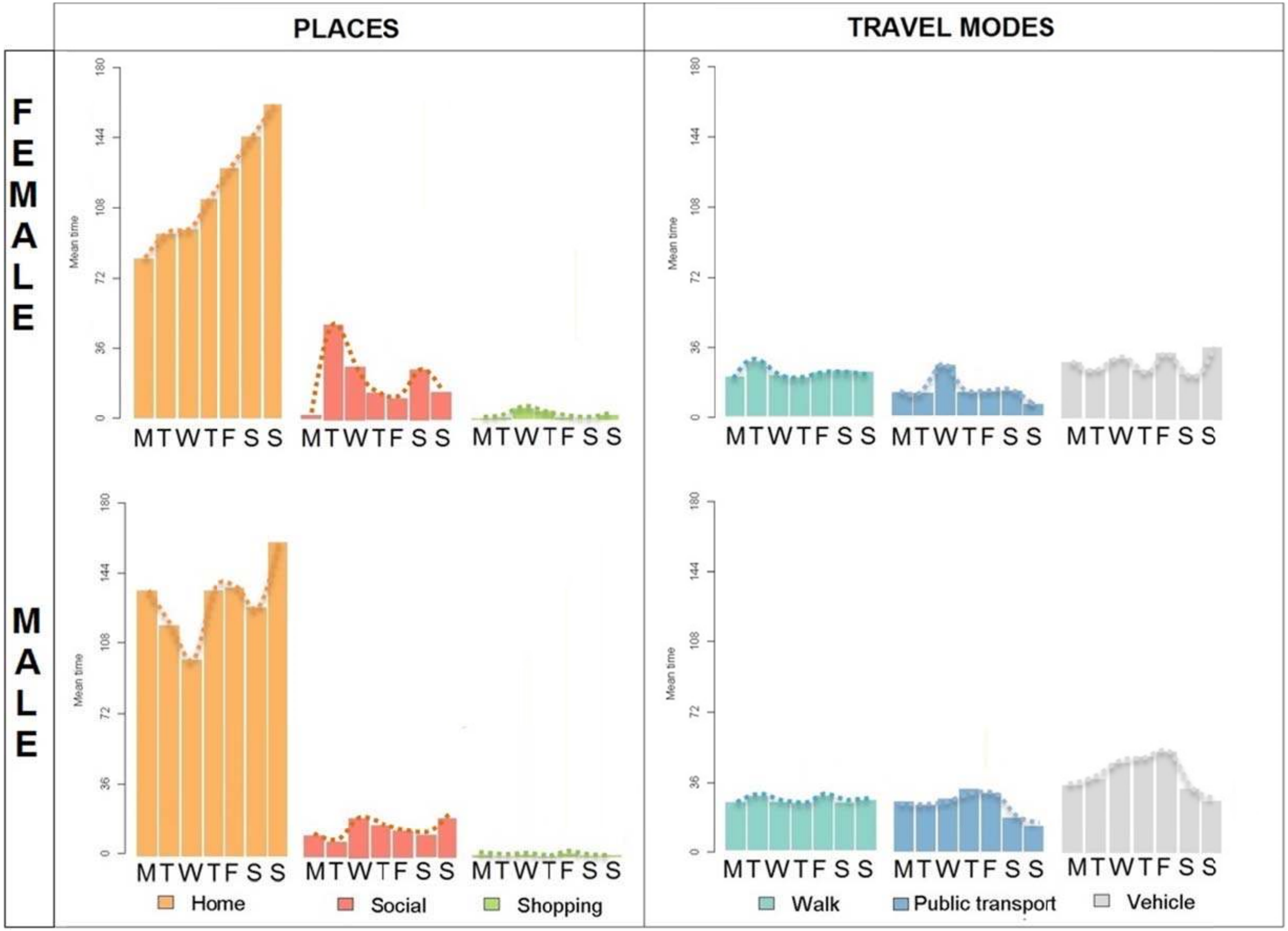

Figure 6 - Average time expenditure and trend over the week (dashed lines) by movement mode, day of the week and gender. The first row refers to females (22 ≤ n ≤ 33) and the second row to males (54 ≤ n ≤ 70).

3.2. Context-Aware Similarity Analysis (CASA)

3.2.1. Cluster analysis & typology

Overall the CHI was higher on weekdays for all weather variables for travel modes and places (Figure 7) indicating the existence of a clear division amongst behavioural groups and a more homogenous movement behaviour within these groups regarding the weather effects in comparison to the weekend. This could also be reflective of the higher diversity of activities during the weekends shown by the higher EI; more variety might lead to lower separability between groups and higher within group distances making more difficult to identify group's responses to weather. This usually results in a higher optimum number of cluster, as seen on the weekend chart (Figure 7 right).

On weekdays the CHI followed a similar pattern for travel modes and places for all weather variables but relative humidity, for which the index was about two times higher for travel modes (Figure 7 left). This indicates a stronger distinction between the two behavioural groups regarding travel modes and relative humidity on weekdays, which might be related to people using relative humidity as a proxy for rainfall to plan their journeys. Relative humidity, rain and comfort showed the highest discernibility for travel mode differences during the week, while for places the highest discernibility was associated with temperature, rain and comfort during the weekend.

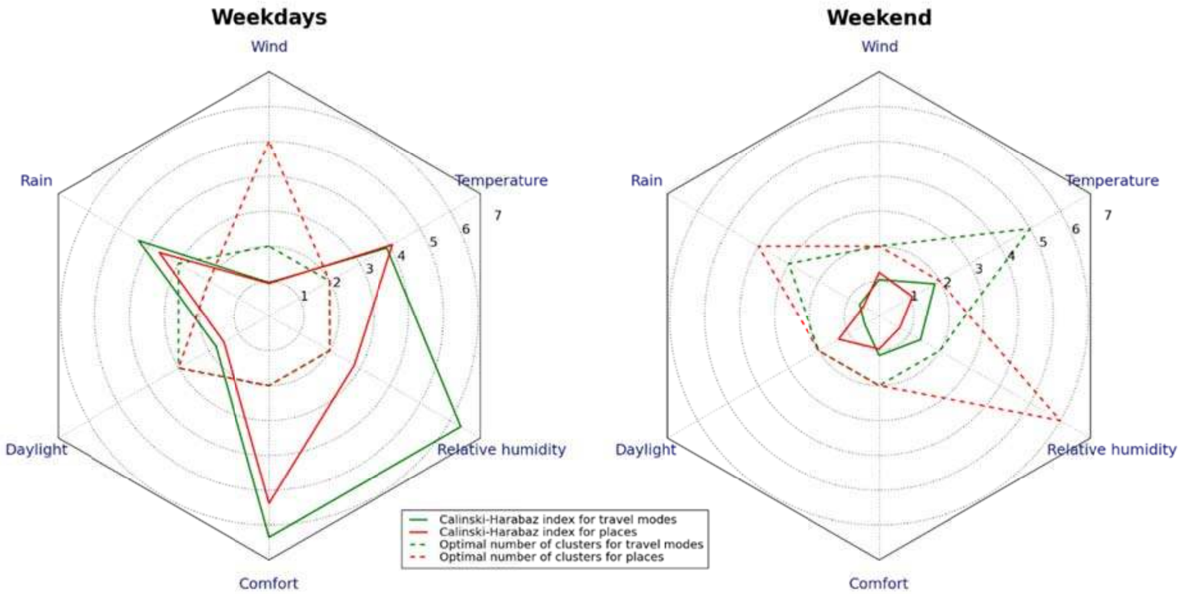

Figure 7 - Calinski-Harabaz index (solid line) and optimal number of clusters (dashed line) for MCSA performed on weather variables for travel modes (green) and places (red) on weekdays and on weekend. The reported CHI is divide by ten and refers to the number of clusters used to split the sequences.

Next, we present some of the more interesting findings while the remaining set of results can be found in the appendix. The typologies are specific for each variable, i.e., Type 1 for wind is not the same as type 1 for rainfall. The analysis of shared discrepancy did not show significant correlation between the behavioural clusters and gender or age groups, all SD were lower than 0.01 with non-significant p-values (α = 0.1). Significant values for Levene's (L) and Kruskal-Wallis' tests (K) are reported on the heading of each graph on the pictures by the following symbology: *** for α = 0.001, ** for α = 0.01, * for α = 0.5, . for α = 0.1.

3.2.2. Wind

Figure 8 shows the clusters for MCSA on wind on weekdays (Figure 8A) and weekends (Figure 8B). The top boxplot shows the distribution of the GPS active time spent under wind blowing from each direction and the middle one shows the distribution of the GPS active time spent under different wind intensities. Both boxplot panels are divided into Type 1 and Type 2, which refer to the two clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the wind conditions encountered within those groups. There were no significant differences on the average time spent on different travel modes on weekdays under different wind conditions (Figure 8A), on the weekend however we found significant differences on the average time expenditure in public transport and vehicle (Figure 8B). CASA clustering showed a significantly lower use of public transportation with concurrent increase on the use of vehicles under more windy conditions coming from North-East, North-West and South-West (Type 2).

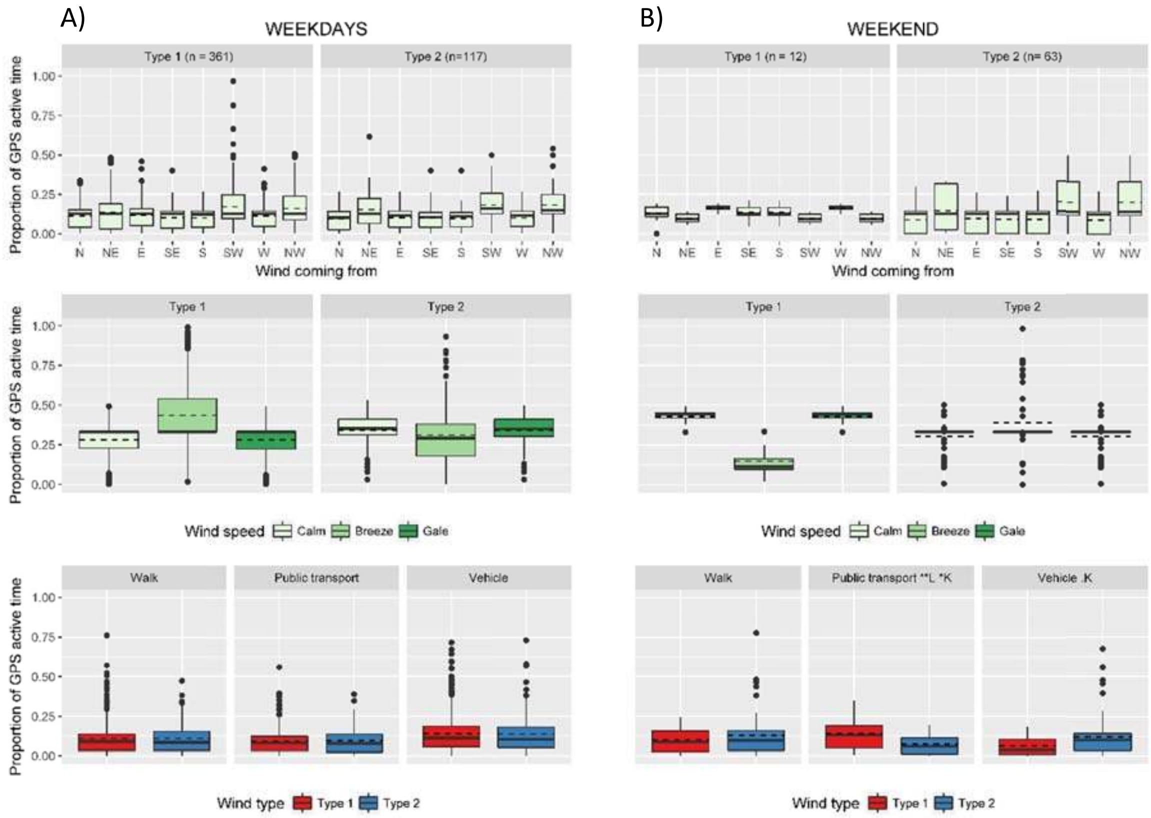

Figure 8 – Clusters for MCSA on wind speed, wind direction and travel modes on weekdays (A) and weekends (B). The four top panels describe the wind conditions within each cluster (Types) and the respective proportions of GPS active time spent under the classes of wind speed and direction. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on each travel mode by the wind types described on the panels above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

Figure 9 shows the clusters for MCSA on wind on weekdays (Figure 9A) and weekends (Figure 9B). The top boxplot shows the distribution of the GPS active time spent under wind blowing from each direction and the middle one shows the distribution of the GPS active time spent under different wind intensities. Both boxplot panels are divided into five types on weekdays and two types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different activities is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on activities, while the remaining panels describe the wind conditions encountered within those groups. For places, there were significant differences in the average time expenditure at home and shopping during the week, and in socialising on the weekend (Figure 9). There were five clusters based on wind during weekdays, but Type 1, Type 3 and Type 4 are very similar in terms of time spent at places and they do not show any pattern in terms of wind direction and strength. We are not able to draw conclusions about Type 5 because of its small number of participants (n = 7); Type 2 however, showed a lower proportion of time spent at home with concurrent increase of time spent shopping under more windy conditions. CASA clustering on weekend showed a significant decrease on the proportional time spent socialising under more windy conditions coming from NE, NW and SW (Type 2). Whereas weekend Type 1 does not show any prevailing direction and its strength alternates between calm and gale for around 88% of the time.

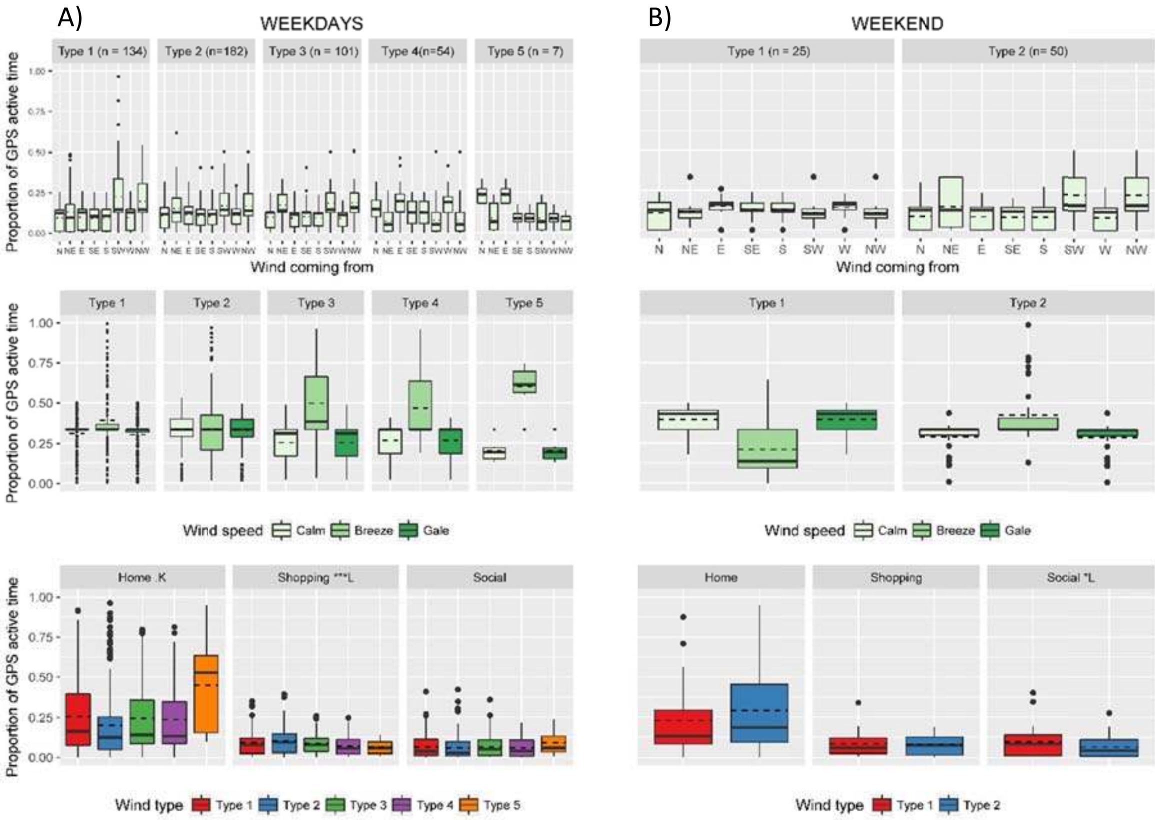

Figure 9 – Clusters for MCSA on wind speed, wind direction and places on weekdays (A) and weekends (B). The four top panels describe wind conditions within each cluster (Types) and respective proportions of GPS active time spent under the classes of wind speed and direction. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on places by the wind types described on the panels above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

3.2.3. Rain

Figure 10 shows the clusters for MCSA on rainfall on weekdays (Figure 10A) and weekends (Figure 10B). The top boxplot shows the distribution of the GPS active time spent under different rainfall intensities. The boxplot panel is divided into three types on weekdays and four types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the rainfall conditions encountered within those groups. There were no significant differences in the average time expenditure for different travel modes on weekends under different rain conditions (Figure 10), on weekdays however we found significant differences in the average time spent in public transport (Figure 10). CASA clustering showed that in comparison to more drier conditions (Type 1 and Type 2), public transport is significantly less used under heavy rainfall (Type 3) with a concurrent, but not statistically significant, increase on the use of vehicles and decrease on walking.

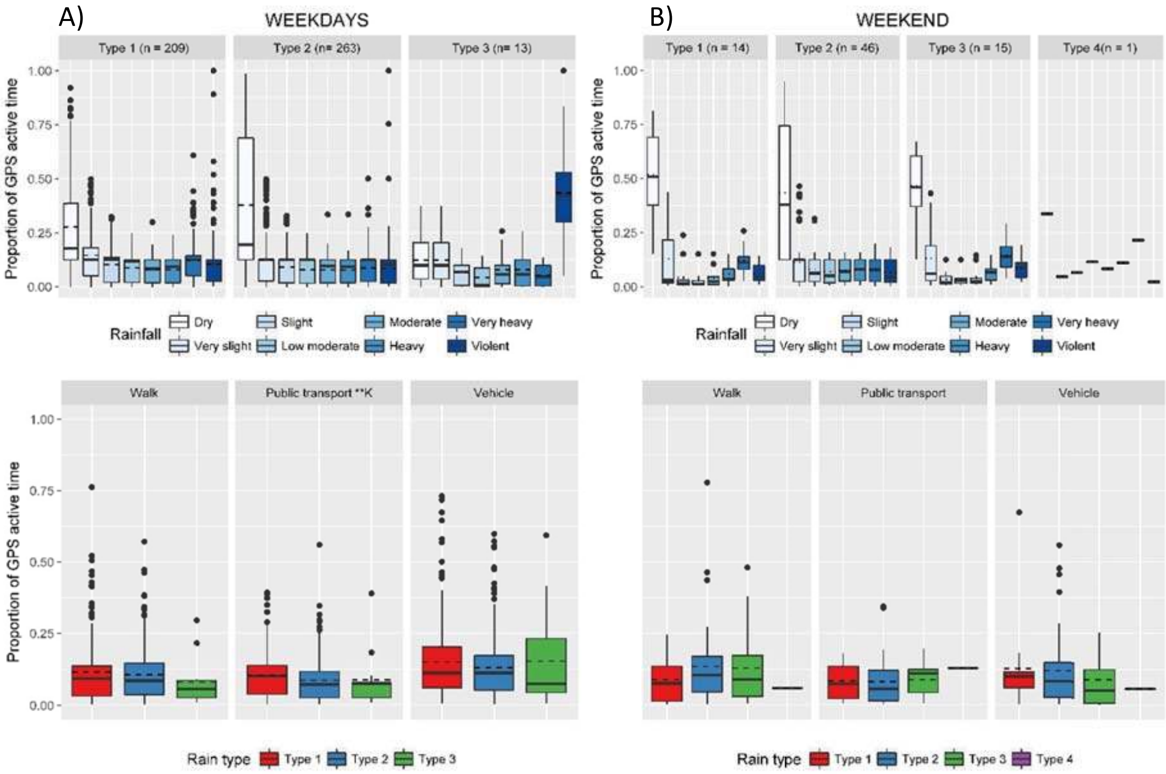

Figure 10 – Clusters for MCSA on rain and travel modes on weekdays (A) and weekends (B). The two top panels describe the rain conditions within each cluster (Types) and the respective proportions of GPS active time spent under the rainfall classes. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on each travel mode by the rain type described on the panel above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

Figure 11 shows the clusters for MCSA on rainfall on weekdays (Figure 11A) and weekends (Figure 11B). The top boxplot shows the distribution of the GPS active time spent under different rainfall intensities. The boxplot panel is divided into three types on weekdays and six types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different activities is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on activities, while the remaining panels describe the rainfall conditions encountered within those groups. The only significant difference for places was on the average time expenditure at home on weekends and weekdays under different rain conditions (Figure 11). On weekdays we found one cluster with predominantly dry conditions (Type 1), a second cluster with dry conditions but an even distribution of time amongst the other rain states (Type 2) and a third cluster with violent rain (Type 3). As expected, people spend more time at home under violent rain (Type 3), but surprisingly people also spend more time at home under predominantly dry conditions (Type 1) in comparison when there is a mix of dry and different rain conditions (Type 2).

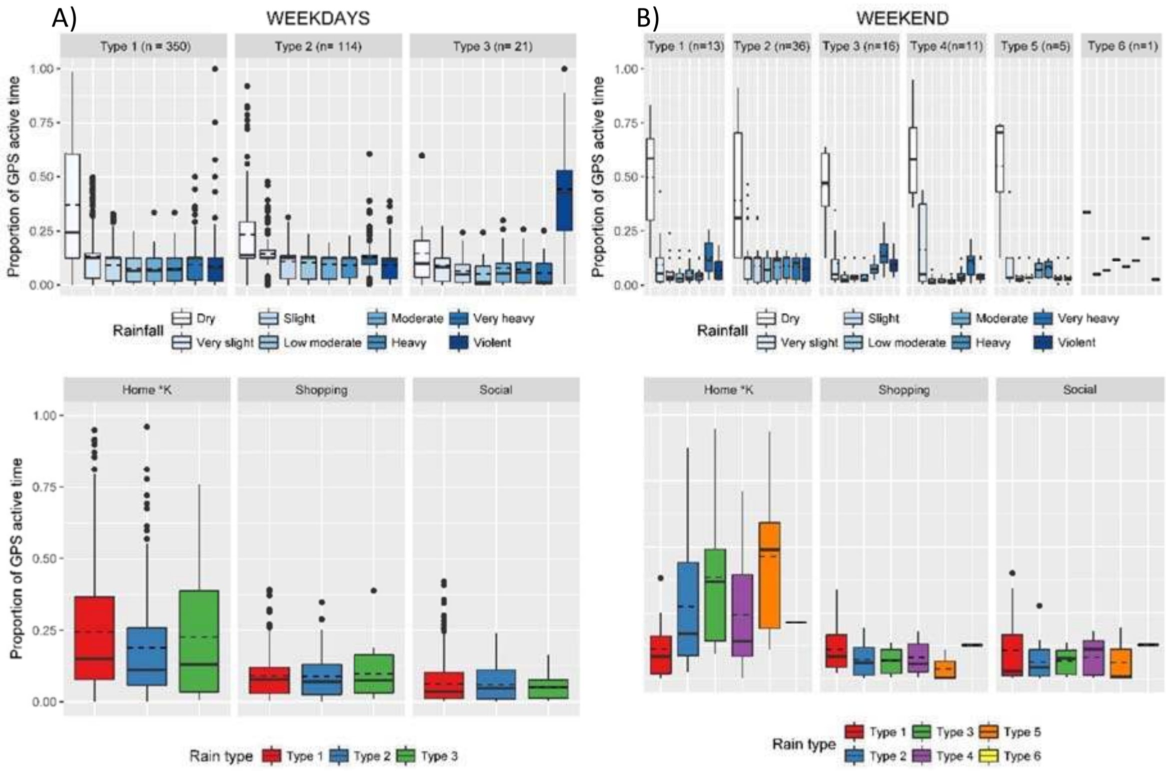

Figure 11 – Clusters for MCSA on rain and places on weekdays (A) and weekends (B). The two top panels describe the rain conditions within each cluster (Types) and the respective proportions of GPS active time spent under the rainfall classes. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on places by the rain type described on the panel above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

On weekends, Type 5 and Type 6 have such a sparse number of members that we considered them outliers. Under heavy rainfall (Type 3) there is a higher average time expenditure at home, while less time is spent at home under predominantly dry conditions, even with the remaining time being almost evenly distributed amongst the other rain conditions (Type 2 and Type 4); the driest conditions (Type 1) showed the higher time expenditure at home.

3.2.4. Daylight

Figure 12 shows the clusters for MCSA on daylight on weekdays (Figure 12A) and weekends (Figure 12B). The top boxplot shows the distribution of the GPS active time spent under different light conditions. The boxplot panel is divided into three types on weekdays and two types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the daylight conditions encountered within those groups. There were no significant differences on the average time expenditure for different travel modes on weekends nor on weekdays under different daylight conditions (Figure 12).

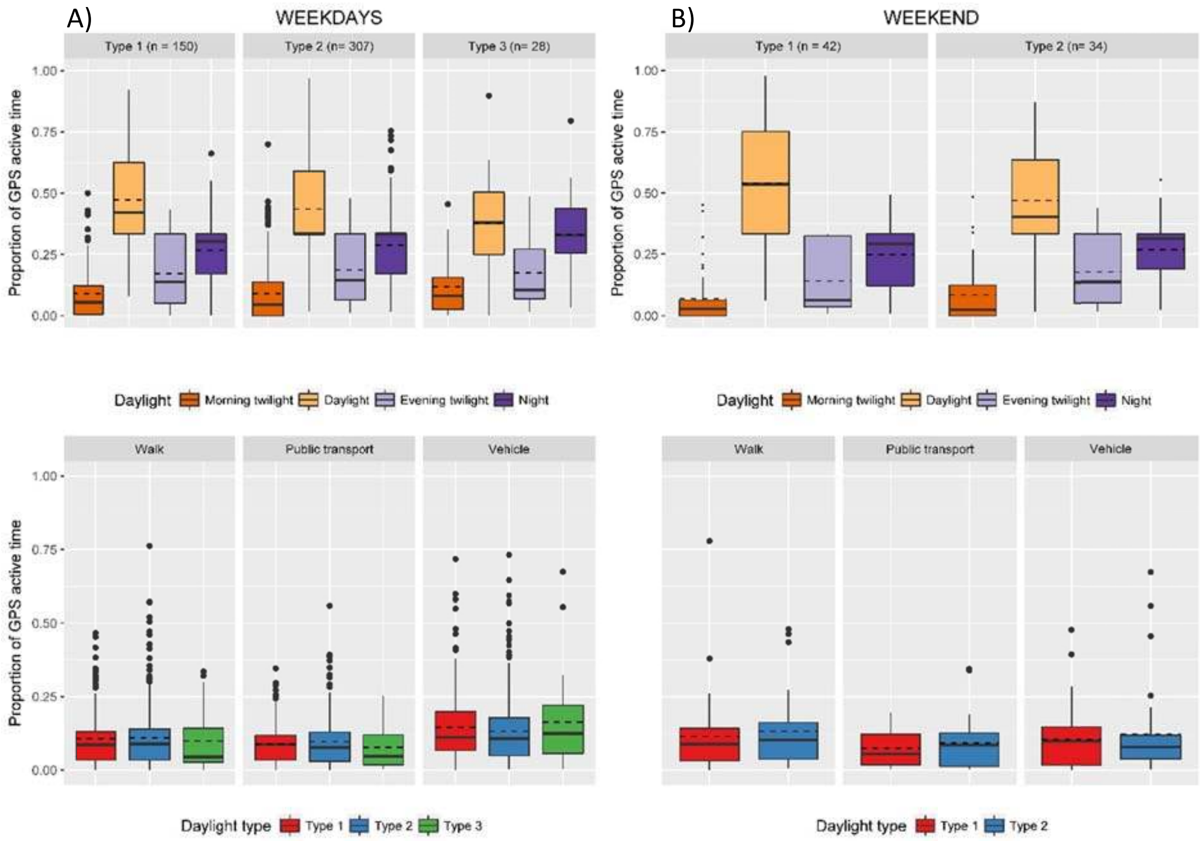

Figure 12 – Clusters for MCSA on daylight and travel modes on weekdays (A) and weekends (B). The two top panels describe daylight conditions within each cluster (Types) and respective proportions of GPS active time spent under the classes of daylight. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on each travel mode by the daylight type described above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

Despite not being statistically significant, less daylight time resulted in less time walking (Type 3) compared to more time walking under more daylight hours (Type 1 and Type 2). A similar decrease is observed on the time expenditure in public transport, with a concurrent increase on time expenditure in vehicles. This trend reverses on weekends, in which walking and public transport are more prominent than the use of vehicle in a group exposed to more night hours (Type 2), while the use of vehicles prevails in a group with more daylight hours (Type 1).

Figure 13 shows the clusters for MCSA daylight on weekdays (Figure 13A) and weekends (Figure 13B). The top boxplot shows the distribution of the GPS active time spent under different light conditions. The boxplot panel is divided into three types on weekdays and two types on weekends, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different activities is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on activities, while the remaining panels describe the daylight conditions encountered within those groups. The analysis for daylight and places, was significant for all places both on weekend and weekdays. There are 3 daylight types on weekdays, Type 1 has more daylight, Type 2 is the one with more night time and Type 3 is an almost even mix of day and night (Figure 13). Type 1 has less time spent at home than Type 2, however we are unsure why the time expenditure at home is the lowest for Type 3. There is less shopping and socialising in the group with more night hours (Type 2). On weekends (Figure 13 B). more time is spent at home under brighter conditions (Type 2), while under lower light conditions (Type 1) more time is spent shopping and socialising.

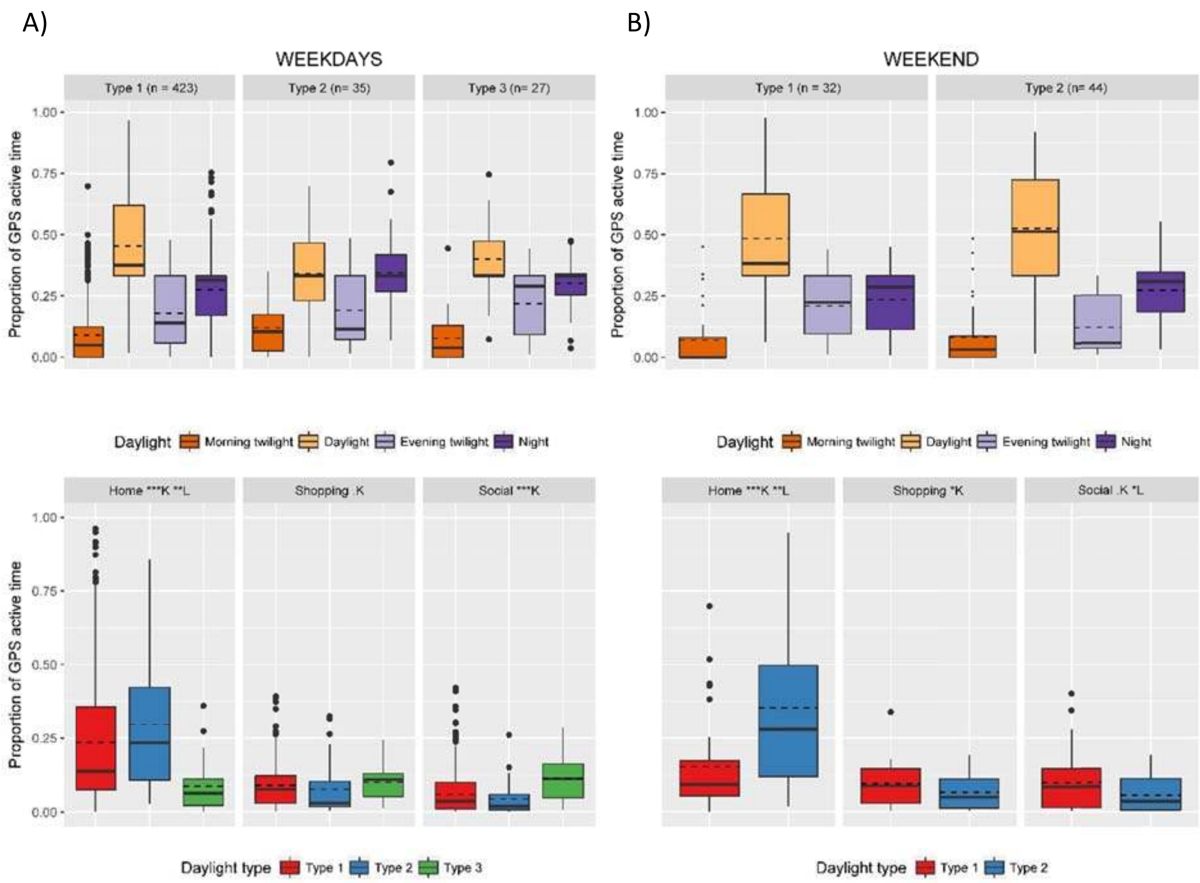

Figure 13 – Clusters for MCSA on daylight and places on weekdays (A) and weekends (B). The two top panels describe daylight conditions within each cluster (Types) and respective proportions of GPS active time spent under daylight classes. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on places by the daylight type described above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

4. Conclusions and discussion

The recent widespread availability and quality of geospatial data on movement and context presents opportunities for developing new methods to understand the interactions between movement behaviour and environment. We were interested on how weather affects human movement, in particular the choice of travel mode and time spent on activities. Our methodology was efficient in identifying groups of specific behaviour under certain weather conditions, and it can be expanded to other types of movement and contextual data. We investigated the impact of wind (strength and direction), rainfall, daylight, comfort, relative humidity and temperature, on the proportion of GPS active time spent on travel modes (walk, public transport, vehicles) and places (home, shopping, socialising). Differences were observed between the time expenditure on different travel modes and places across the day, week and between genders. The analysis of the entropy index (EI) showed a high diversity of movement modes in the early morning on weekdays and weekends, with a positive shift of three hours on weekends. Horanont et al. (2013) found the same entropy pattern, despite analysing weekdays and weekends together, when using GPS traces from mobile phone to explore the effects of weather on daily routine. We found that during weekdays there is a drop with subsequent rise on the EI, which does not exist on weekends because the EI is higher from 10 am throughout the afternoon. Horanont et al. (2013) found a very similar variation for specific extreme weather conditions, according to meteorological information provided by the authors, which the authors attributed to the weather conditions. However, we believe it is related to similar differences to the ones we found between weekdays and weekend, and that it is more likely that the extreme weather events reported by Horanont et al. (2013) took place on a weekend. Similarly to Ryan et al. (2010), we found that people have more varied activities on weekends, which was shown by the highest EI on Sunday and Saturday. This happens because people have more scope for freedom of action on weekends, in contrast to the external controls imposed on weekdays by work and school (Ryan et al., 2010).

The average time spent (AVTS) walking did not change substantially across weekdays and between genders (Figure 6). The AVTS at home varied throughout the week, being the highest on Sunday and lowest on Wednesday for both genders (dashed orange lines on Figure 6). The low values on Wednesday might be related to the higher average time spend socializing in comparison to other days of the week (dashed red lines on Figure 6). Moreover, women seem to spend more time socializing and to concentrate social activities on Tuesdays, Wednesdays and Saturdays; while men socialise very little on Tuesdays and keep a steady, but lower than women, average from Wednesday to Monday.

Similarly to Stover et al. 2012, we found that the wind strength and direction exerted considerable influence on weekends on the use of public transport and vehicles, which is possibly related to traffic restrictions at the Tay and Forth bridges (See dashed ellipses in Figure 1) under high winds, which are more likely to come from NW, SW and NE in the Central Belt of Scotland. There were at least ten occasions during our data collection during which the bridges were either closed or had restrictions on the type of vehicle and speed limits because of high winds (Traffic Scotland '@trafficscotland' 2017). It is possible that these restrictions are reflected in our findings during these windy periods, since the participants in our study were mostly commuters from Fife to Edinburgh or Dundee (Sila-Nowicka et al. 2016) and therefore typically have to cross one of these two bridges daily.

As opposed to what Guo et al. (2007) found in Chicago, we found that rain during the weekend has no key role on travel modes, but heavy rain decreases the use of public transport during the week. This could be explained by the fact that discretionary passengers are more affected by rain than commuters (Changnon, 1996), i.e., people are obliged to go out for their daily duties on weekdays and therefore might adapt their travel modes, while on weekends they can opt to stay at home under heavy rain. In addition, similarly to what Chen et al. (2017) found when studying the impact of rainfall on taxi use, we also found a trend of more vehicular use under rainy weather, and less walking in heavier rain.

We found that daylight length seems to factor into mobility decisions differently on weekends in comparison to weekdays. During the week, less daylight hours were linked to less walking and less public transport use, but more vehicular use; on weekends the same daylight conditions resulted in the opposite pattern It is not clear why this may be. In addition, daylight seems to play a major role on time expenditure at certain places; weekdays with more dark hours are more likely to be spent at home, while more time is spent at home on weekends under more daylight hours. Temperature increase seems to have a positive effect on walking (Cools, Moons, Creemers, & Wets, 2010; Tucker & Gilliland, 2007), which makes sense in Scotland because the temperate and oceanic climate gives people more opportunities for outdoor activities during summer. It is likely that in places with more tropical climates the temperature effect would be different, in the USA for example, areas with a more tropical weather showed a decline in physical activity on hotter months, while areas with cold weather showed an increase on warmer months (Tucker & Gilliland, 2007).

The application of multi-channel sequence analysis on semantic trajectories was efficient for identification of movement patterns. Even though our method does not use the exact coordinates, the multi-channel sequences keep the spatial component through places and travel modes, which allows us to link movement patterns to environmental conditions and identify responses. Our methodology works both with categorical and numerical contextual data, considers the change of context between two timestamps, is able to handle multiple contextual variables and their interactions at once, and can deal with contextual data in any the form. These capabilities make it more able to deal with complex contextual situations than previous methodologies, such as those established by Sharif & Alesheikh (2017b), De Groeve et al. (2016) and Buchin et al (2014).

MCSA clusters are useful for simplifying the increasingly large and complex tracking datasets, the creation of typologies allows the generalization and reduction of thousands of trajectories to a few representative trajectories. In addition, MCSA can help with the recent increasing demand for Context-Aware methods, as it is able to perform similarity analysis taking context into account and also allows for visualization of movement patterns and contextual variables simultaneously along the time axis. Another advantage here is that the time units are flexible, i.e., the sequences can be arranged at daily, weekly, monthly or hourly scale, which allows for multi-scale detection of movement patterns.

To summarise, multi-channel sequence analysis represents a new analysis tool for movement data where contextual information can now be readily combined with detailed tracking datasets. The main advantage of this approach is that it also is possible to consider as many channels (variables) as desired at once. It is common in movement research to simultaneously consider multiple environmental variables, which makes MCSA particularly relevant for studying human mobility, traffic, transportation and wildlife ecology; areas in which movement behaviour may be contextualised by other dynamic environmental variables such as air temperature, vegetation indices, humidity, wind speed, air pollution and snow coverage. We believe that MCSA can help performing Context-Aware Similarity Analysis (CASA), which improves our understanding of how movement is affected by the combination of multiple contextual variables. In addition, MCSA is a good approach to summarise large movement dataset into clusters expressing specific typologies, i.e., a group of similar movement patterns.

Acknowledgments

The authors would like to acknowledge the Brazilian government and SWB (Science Without Borders) - CAPES (Coordination for the Improvement of Higher Education Personnel) for the financial support during the development of this study. United Kingdom Meteorological Office (UKMO) provided the MIDAS: UK Hourly Weather Observation Data and the Rain Radar Data from the NIMROD system, both via NERC Data Centre and under the following licence:

http://licences.ceda.ac.uk/image/data_access_condition/ukmo_agreement.pdf .

References

Abbott, A. (1995). Sequence Analysis: New Methods for Old Ideas. *Source: Annual Review of Sociology Annu. Rev. Sociol*, *21*(21). Retrieved from http://www.jstor.org/stable/2083405

Abbott, A., & Tsay, A. (2000). Sequence Analysis and Optimal Matching Methods in Sociology. *Sociological Methods & Research*, *29*(1), 3–33. http://doi.org/10.1177/0049124100029001001

Ahmed, K. I. (2012). Acoustic data optimisation for seabed mapping with visual and computational data mining.

Anderson, C. A. (2001). Heat and violence. *Current Directions in Psychological Science*, *10*(1), 33–38. Retrieved from http://www2.psych.ubc.ca/~schaller/308Readings/Anderson2001.pdf

Andrienko, G., Andrienko, N., & Heurich, M. (2011). An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, *25*(9), 1347–1370. http://doi.org/10.1080/13658816.2011.556120

Ayoade, J. . (1986). *Introdução à climatologia para os trópicos*. São Paulo: Difel.

Batagelj, V. (1988). Generalized Ward and Related Clustering Problems. In H. H. Bock (Ed.), *n Classification and related methods of data analysis* (pp. 67–74). Amsterdam. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.5954&rep=rep1&type=pdf

Beecham, R., Wood, J., & Bowerman, A. (2014). Studying commuting behaviours using collaborative visual analytics. *Computers, Environment and Urban Systems*, *47*, 5–15. http://doi.org/10.1016/j.compenvurbsys.2013.10.007

Billari, F. C. (2001). Sequence Analysis in Demographic Research. *Special Issue on Longitudinal Methodology Canadian Studies in Population*, *28*(2), 439–458. Retrieved from http://www.canpopsoc.ca/CanPopSoc/assets/File/publications/journal/CSPv28n2p439.pdf

Brum-Bastos, V., Long, J., & Demšar, U. (2016). Dynamic trajectory annotation for integrating environmental and movement data. In *Visually-supported Computational Movement Analysis Workshop - AGILE 2016*. Helsinki. Retrieved from http://viz.icaci.org/vcma2016/wp-content/uploads/2016/07/vcma_brumbastos-paper.pdf

Buchin, M., Dodge, S., & Speckmann, B. (2014). Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science*, *9*(9), 101–124. http://doi.org/10.5311/JOSIS.2014.9.179

Cabanac, M. (1971). Physiological role of pleasure. *Science (New York, N.Y.)*, *173*(4002), 1103–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5098954

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, *3*(1), 1–27. http://doi.org/10.1080/03610927408827101

Carlsmith, J. M., & Anderson, C. A. (1979). Ambient Temperature and the Occurrence of Collective Violence: ANew Analysis. *Journal of Personality and Social Psychology*, *37*(3), 337–344. Retrieved from https://public.psych.iastate.edu/caa/abstracts/1979-1984/79CA.PDF

Changnon, S. A. (1996). Effects of summer precipitation on urban transportation. *Climatic Change*, *32*(4), 481–494. http://doi.org/10.1007/BF00140357

Chen, D., Zhang, Y., Gao, L., Geng, N., & Li, X. (2017). The impact of rainfall on the temporal and spatial distribution of taxi passengers. *PLOS ONE*, *12*(9). http://doi.org/10.1371/journal.pone.0183574

Connolly, M. (2008). Here Comes the Rain Again: Weather and the Intertemporal Substitution of Leisure. *Source Journal of Labor Economics Journal of Labor Economics*, *26*(1), 73–100. http://doi.org/10.1086/522067

Cools, M., Moons, E., Creemers, L., & Wets, G. (2010). Changes in Travel Behavior in Response to Weather Conditions: Whether Type of Weather and Trip Purpose Matter? *Journal of the Transportation Research Board*, *2157*, 22–28. http://doi.org/http://dx.doi.org/10.3141/2157-03

de Freitas, C. R. (1990). Recreation climate assesment. *International Journal of Climatology*, *10586*(551), 89–103. Retrieved from https://s3.amazonaws.com/academia.edu.documents/40659072/DeFreitas_1990_IJC_Rec_Climate_Assessment_1990.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1505291213&Signature=%2FiR2Mu%2F2jB%2FHEtvqxR5HwoGAJok%3D&response-content-disposition=inline%3B file

de Freitas, C. R. (2003). Tourism climatology: evaluating environmental information for decision making and business planning in the recreation and tourism sector. *International Journal of Biometeorology*, *48*(1), 45–54. http://doi.org/10.1007/s00484-003-0177-z

de Freitas, C. R. (2015). Weather and place-based human behavior: recreational preferences and sensitivity. *International Journal of Biometeorology*, *59*(1), 55–63. http://doi.org/10.1007/s00484-

014-0824-6

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O., & Cagnacci, F. (2016). Extracting spatio-temporal patterns in animal trajectories: An ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, *7*(3), 369–379. http://doi.org/10.1111/2041-210X.12453

de Montigny, L., Ling, R., & Zacharias, J. (2012). The Effects of Weather on Walking Rates in Nine Cities. *Environment and Behavior*, *44*(6), 821–840. http://doi.org/10.1177/0013916511409033

Demsar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., … Weibel, R. (2015). Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*, *3*(1). http://doi.org/10.1186/s40462-015-0032-y

Derrick Sewell, W. R., Kates, R. W., & Philips, L. E. (1968). Human responde to weather and climate : geographical contributions. *The Geoographical Review*, *LVIII*(2), 262–280.

Dodge, S., Bohrer, G., Weinzierl, R., Davidson, S. C., Kays, R., Douglas, D., … Wikelski, M. (2013). The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data. *Movement Ecology*, *1*(1), 3. http://doi.org/10.1186/2051-3933-1-3

Dodge, S., Weibel, R., Ahearn, S. C., Buchin, M., & Miller, J. A. (2016). Analysis of movement data. *International Journal of Geographical Information Science*, *30*(5), 825–834. http://doi.org/10.1080/13658816.2015.1132424

Dunne, S., & Ghosh, B. (2013). Weather Adaptive Traffic Prediction Using Neurowavelet Models. *IEEE Transactions on Intelligent Transportation Systems*, *14*(1), 370–379. http://doi.org/10.1109/TITS.2012.2225049

Gabadinho, A., Ritschard, G., Studer, M., & Müller, N. S. (2009). Mining Sequence Data in R with the TraMineR package: A user's guide. University of Geneva: Department of Econometrics and Laboratory of Demography. Retrieved from https://cran.r-project.org/web/packages/TraMineR/TraMineR.pdf

Gauthier, J.-A., Widmer, E. D., Bucher, P., & Notredame, C. (2010). 1. Multichannel Sequence Analysis Applied to Social Science Data. *Sociological Methodology*, *40*(1), 1–38. http://doi.org/10.1111/j.1467-9531.2010.01227.x

Givoni, B. (1974). Architectural and urban planning in relation to weather and climate. *Progress in Biometeorology. Division A: Progress in Human Biometeorology*, *1*(1A), 183–93, 632. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4438650

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*, *36*(2), 131–139. http://doi.org/10.1016/j.compenvurbsys.2011.05.003

Guo, Z., Wilson, N., & Rahbee, A. (2007). Impact of Weather on Transit Ridership in Chicago, Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, *2034*, 3–10. http://doi.org/10.3141/2034-01

Horanont, T., Phithakkitnukoon, S., Leong, T. W., Sekimoto, Y., & Shibasaki, R. (2013). Weather effects on the patterns of people's everyday activities: A study using GPS traces of mobile phone users. *PLoS ONE*, *8*(12), 1–14. http://doi.org/10.1371/journal.pone.0081153

Horowitz, S. (2002). Weather–Health Links: Epidemiologic Research, Clinical Observations, and Patient Perceptions. *Alternative & Complementary Therapies*, 34–40. Retrieved from http://www.sld.cu/galerias/pdf/p34_s.pdf

Idury, R. M., & Waterman, M. S. (1995). A New Algorithm for DNA Sequence Assembly. *Journal of Computational Biology*, *2*(2), 291–306. http://doi.org/10.1089/cmb.1995.2.291

Jenkins, G., Perry, M., & Prior, J. (2009). *The climate of the United Kingdom and recent trends*. Met Office Hadley Centre, Exeter, UK. Retrieved from http://www.ukcip.org.uk/wp-content/PDFs/UKCP09_Trends.pdf

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2018). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7. Comprehensive R Archive Network (CRAN). Retrieved from https://cran.r-project.org/web/packages/cluster/index.html

Manu, S., Shukla, Y., Rawal, R., Thomas, L. E., & de Dear, R. (2016). Field studies of thermal comfort across multiple climate zones for the subcontinent: India Model for Adaptive Comfort (IMAC). *Building and Environment*, *98*, 55–70. http://doi.org/10.1016/j.buildenv.2015.12.019

Meekan, M. G., Duarte, C. M., Fernández-Gracia, J., Thums, M., Sequeira, A. M. M., Harcourt, R., & Eguíluz, V. M. (2017). The Ecology of Human Mobility. *Trends in Ecology and Evolution*, *32*(3), 198–210. http://doi.org/10.1016/j.tree.2016.12.006

Meijles, E. W., de Bakker, M., Groote, P. D., & Barske, R. (2014). Analysing hiker movement patterns using GPS data: Implications for park management. *Computers, Environment and Urban Systems*, *47*, 44–57. http://doi.org/10.1016/j.compenvurbsys.2013.07.005

Met Office. (2012). Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). NCAS British Atmospheric Data Centre. Retrieved from http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0

MetOffice. (2003). 1 km Resolution UK Composite Rainfall Data from the Met Office Nimrod System. NCAS British Atmospheric Data Centre. Retrieved from http://catalogue.ceda.ac.uk/uuid/27dd6ffba67f667a18c62de5c3456350

Murtagh, F., & Legendre, P. (2011). Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. http://doi.org/10.1007/s00357-014-9161-z

Nerlich, B., & Jaspal, R. (2014). Images of Extreme Weather: Symbolising Human Responses to Climate Change. *Science as Culture*, *23*(2), 253–276. http://doi.org/10.1080/09505431.2013.846311

Ng, E. (2012). Towards planning and practical understanding of the need for meteorological and climatic information in the design of high-density cities: A case-based study of Hong Kong. *International Journal of Climatology*, *32*(4), 582–598. http://doi.org/10.1002/joc.2292

Oshan, T., Vandrol, J., Stewart Fotheringham, A., Demsar, U., & Pozdnoukhov AlexeiPozdnoukhov, A. (2014). *GEOCROWD - Creating a Geospatial Knowledge World : Spatiotemporal patterns D1.9*. Retrieved from http://www.geocrowd.eu/wp-content/uploads/2015/01/D1.9_USTAN.pdf

Royal Meteorological Society. (2017). Beaufort Scale. Retrieved September 12, 2017, from https://www.rmets.org/weather-and-climate/observing/beaufort-scale

Ryan, R. M., Bernstein, J., & Warren Brown, K. (2010). Weekends, Work, and Well-Being: Psychological need satisfactions and day of the WeeK effects on mood, vitality, and Physical symptoms. *Journal of Social and Clinical Psychology*, *29*(1), 95–122. Retrieved from https://selfdeterminationtheory.org/SDT/documents/2010_RyanBernsteinBrown_Weekends_JSCP.pdf

Schneider, F. W., Lesko, W. A., & Garret, W. A. (1980). Helping behavior in hot, comfortable, and cold temperatures: A field study. *Environment and Behavior*, *12*(2), 231–241. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/0013916580122007

Sharif, M., & Alesheikh, A. A. (2017a). Context-aware movement analytics: implications, taxonomy, and design framework. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1233. http://doi.org/10.1002/widm.1233

Sharif, M., & Alesheikh, A. A. (2017b). Context-awareness in similarity measures and pattern discoveries of trajectories: a context-based dynamic time warping method. *GIScience and Remote Sensing*, *54*(3), 426–452. http://doi.org/10.1080/15481603.2017.1278644

Shoval, N., & Isaacson, M. (2007). Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American …*, *97*(May 2006), 282–297. http://doi.org/10.1111/j.1467-8306.2007.00536.x

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, *30*(5), 881–906. http://doi.org/10.1080/13658816.2015.1100731

Steadman, R. G. (1994). Norms of apparent temperature in Australia. *Australian Meteorology Magazine*, *43*, 1–16. Retrieved from http://www.bom.gov.au/amm/docs/1994/steadman.pdf

Stover, V. W., Consulting Associates, N., & Edward McCormack, S. D. (2012). The Impact of Weather on Bus Ridership in Pierce County The Impact of Weather on Bus Ridership in Pierce County, Washington. *Journal of Public Transportation*, *15*(1), 95–110. Retrieved from https://www.nctr.usf.edu/wp-content/uploads/2012/04/JPT15.1Stover.pdf

Studer, M., Ritschard, G., Gabadinho, A., & Müller, N. S. (2011). Discrepancy analysis of state sequences. *Sociological Methods and Research*, *40*(3), 471–510. http://doi.org/10.18637/jss.v040.i04

Thakuriah, P. (Vonu), Sila-Nowicka, K., & Paule, J. G. (2016). Sensing Spatiotemporal Patterns in Urban Areas : Analytics and Visualizations Using the Integrated Multimedia City Data. *Built Environment*, (October), 415–429. http://doi.org/10.2148/benv.42.3.415

Traffic Scotland (@trafficscotland). (2017). Twitter advanced search: "cars only" from:trafficscotland since:2013-09-28 until:2014-01-10. Tweet. Retrieved from https://twitter.com/search?l=&q=%22cars only%22 from%3Atrafficscotland since%3A2013-09-28 until%3A2014-01-10&src=typd

Tucker, P., & Gilliland, J. (2007). The effect of season and weather on physical activity: A systematic review. *Public Health*, *121*(12), 909–922. http://doi.org/10.1016/j.puhe.2007.04.009

UK Weather. (2013). Dunfermline monthly Sunset Sunrise times UK. Retrieved September 18, 2017, from http://www.ukweathercams.co.uk/sunrise_sunset_times_by_month.php?id=11662&ts=1505740144

VDI. (2008). VDI 3787, Part I: Environmental Meteorology, Methods for the human biometeorological evaluation of climate and air quality for the urban and regional planning at regional level. Part I: Climate. Berlin: Beuth.

Versichele, M., Neutens, T., Delafontaine, M., & Van de Weghe, N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, *32*(2), 208–220. http://doi.org/10.1016/j.apgeog.2011.05.011

Worfolk, J. B. (1997). Keep Frail Elders Warm! The thermal instabilities of the old have not received sufficient at- tention in basic educational programs. *Geriatric Nursing*, *18*(1), 7–11. Retrieved from http://www.gnjournal.com/article/S0197-4572(97)90123-3/pdf

Zivin, J. G. (2014). Temperature and the Allocation of Time: Implications for Climate Change. *Journal of Labor Economics*, *32*(1). Retrieved from https://www.journals.uchicago.edu/doi/pdfplus/10.1086/671766

APPENDIX

Comfort, humidity and temperature

Figure 14 shows the clusters for MCSA on comfort on weekdays (Figure 14A) and weekends (Figure 14B). The top boxplot shows the distribution of the GPS active time spent under different comfort conditions. The boxplot panel is divided into two types, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the thermal comfort conditions encountered within those groups. The only significant difference for comfort and travel modes happened on weekdays for the average time spent in vehicles under different comfort conditions (Figure 14 Appendix). Slightly uncomfortable conditions (Type 2) were associated with significant higher use of vehicles, less walking and less use of public transport. For the weekend participants were split into one large group and an individual, therefore limiting interpretation. There were no significant differences or meaningful visual patterns from the average time expenditure on different places both on weekdays and weekends under different comfort levels (Figure 15 appendix).

Figure 16 shows the clusters for MCSA on relative humidity on weekdays (Figure 16A) and weekends (Figure 16B). The top boxplot shows the distribution of the GPS active time spent under different relative humidity conditions. The boxplot panel is divided into types, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the meteorological conditions encountered within those groups. The only significant difference for relative humidity and travel modes happened on weekdays on the average time spent walking under different relative humidity (Figure 16 Appendix). It seems that more humid conditions (Type 1) were associated with a significantly higher time expenditure walking. On weekends the time spent in public transport is visually higher when humidity is lower (Type 2).

Figure 17 shows the clusters for MCSA relative humidity on weekdays (Figure 17A) and weekends (Figure 17B). The top boxplot shows the distribution of the GPS active time spent under different relative humidity conditions. The boxplot panel is divided into two types, which refer to the two clusters found by the MCSA analysis and for which the distribution of the GPS active time in different activities is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on activities, while the remaining panels describe the meteorological conditions encountered within those groups. For the analysis on places (Figure 17 Appendix), there were significant differences on time spent at home and socialising on the weekend. Also, when relative humidity is lower (Type 2), more time is spent at home and less time is spent socialising, while it goes the other way around for Type 1, for which the relative humidity is higher. We believe that these patterns are more related to rain than to relative humidity, as higher humidity is closely related to probability of rain.

Figure 18 shows the clusters for MCSA on temperature on weekdays (Figure 18A) and weekends (Figure 18B). The top boxplot shows the distribution of the GPS active time spent under different temperature conditions. The boxplot panel is divided into two types, which refer to the clusters found by the MCSA analysis and for which the distribution of the GPS active time in different travel modes is shown on the boxplot panel at the bottom. This boxplot shows the difference between groups with different distribution of time spent on travel modes, while the remaining panels describe the meteorological conditions encountered within those groups. For temperature and travel modes the only significant differences happened on weekdays for the average time spent walking and on weekends on public transport (Figure 18 Appendix). Higher temperatures on weekdays (Type 1) led to a significant higher time expenditure walking, and slightly less time spent on public transport under temperatures close to the average historical maximum (Type 2). On the other hand, extremely elevated temperatures (Type 1) show more time spent on public transport. Types 3, 4 and 5 had a low number of trajectories (n<9) assigned to them and were considered outliers on which we cannot not draw conclusions.

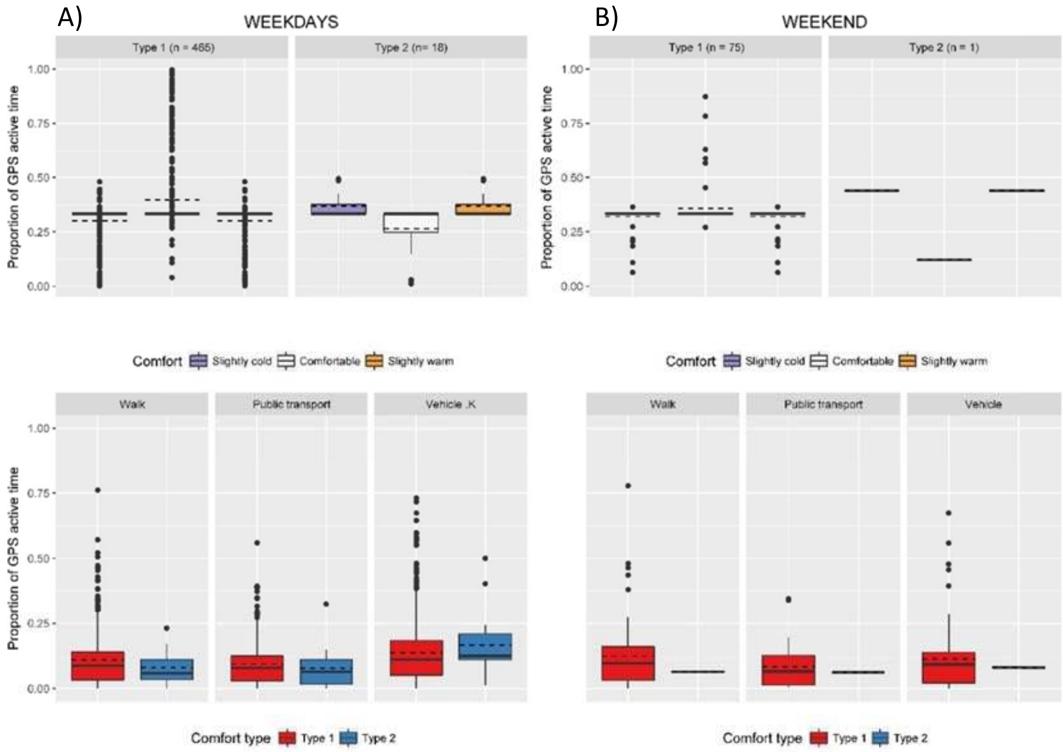

Figure 14 – Clusters for MCSA run on comfort and travel modes on weekdays (A) and weekends (B). The two top panels describe comfort conditions within each cluster (Types) and respective proportions of GPS active time spent under comfort classes. The two panels at the bottom show the distribution of proportional GPS active time spent on each travel mode by the comfort type described above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

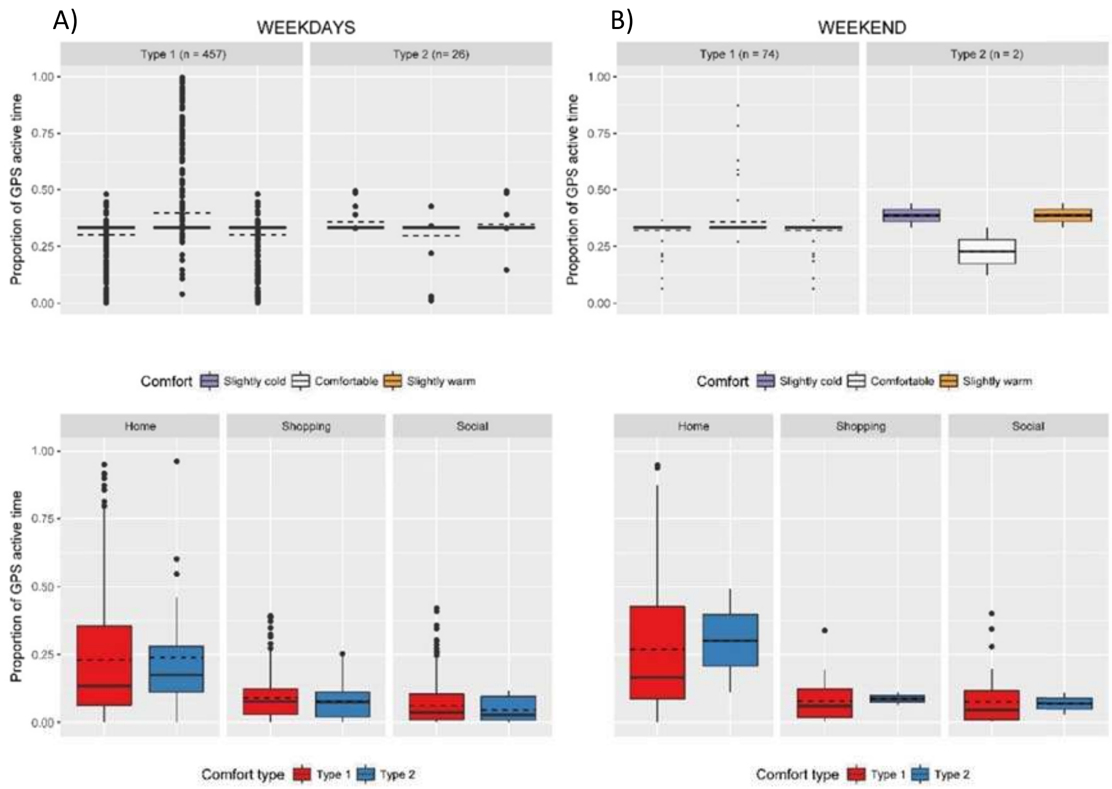

Figure 15 – Clusters for MCSA run on comfort and places on weekdays (A) and weekends (B). The two top panels describe comfort conditions within each cluster (Types) and respective proportions of GPS active time spent under comfort classes. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on places by comfort type described above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

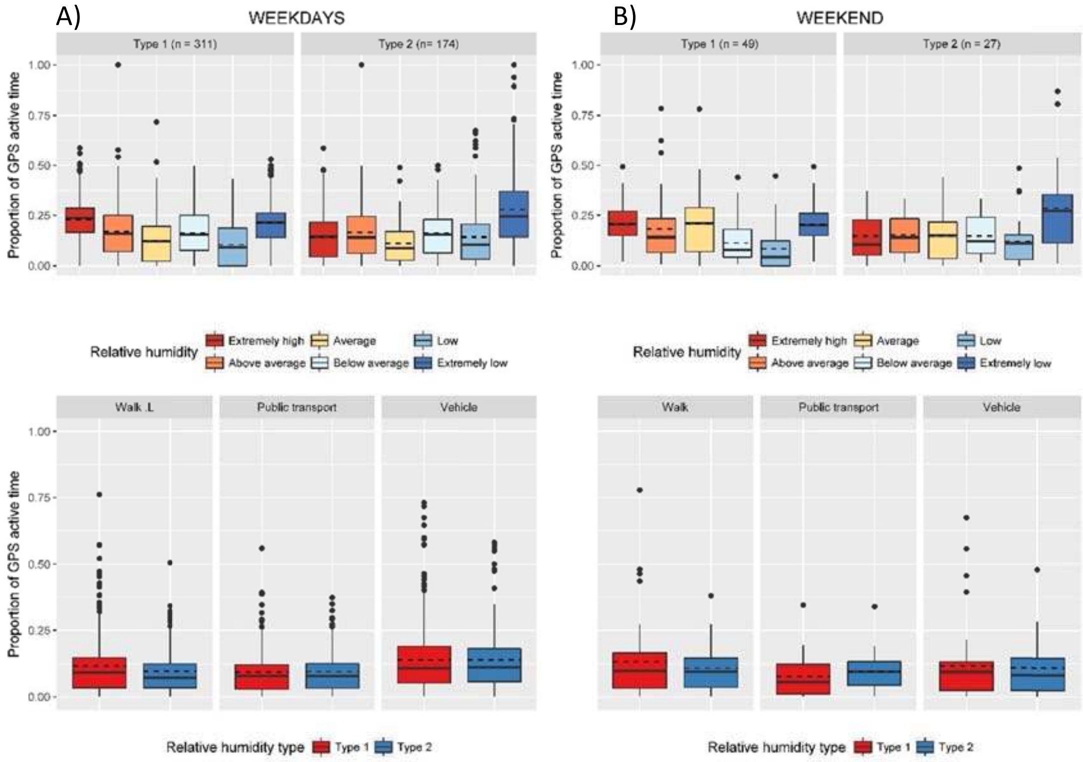

Figure 16 – Clusters for MCSA run on relative humidity and travel modes on weekdays (A) and weekends (B). The two top panels describe relative humidity conditions within each cluster (Types) and respective proportions of GPS active time spent under relative humidity classes. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on each travel mode by relative humidity type described above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

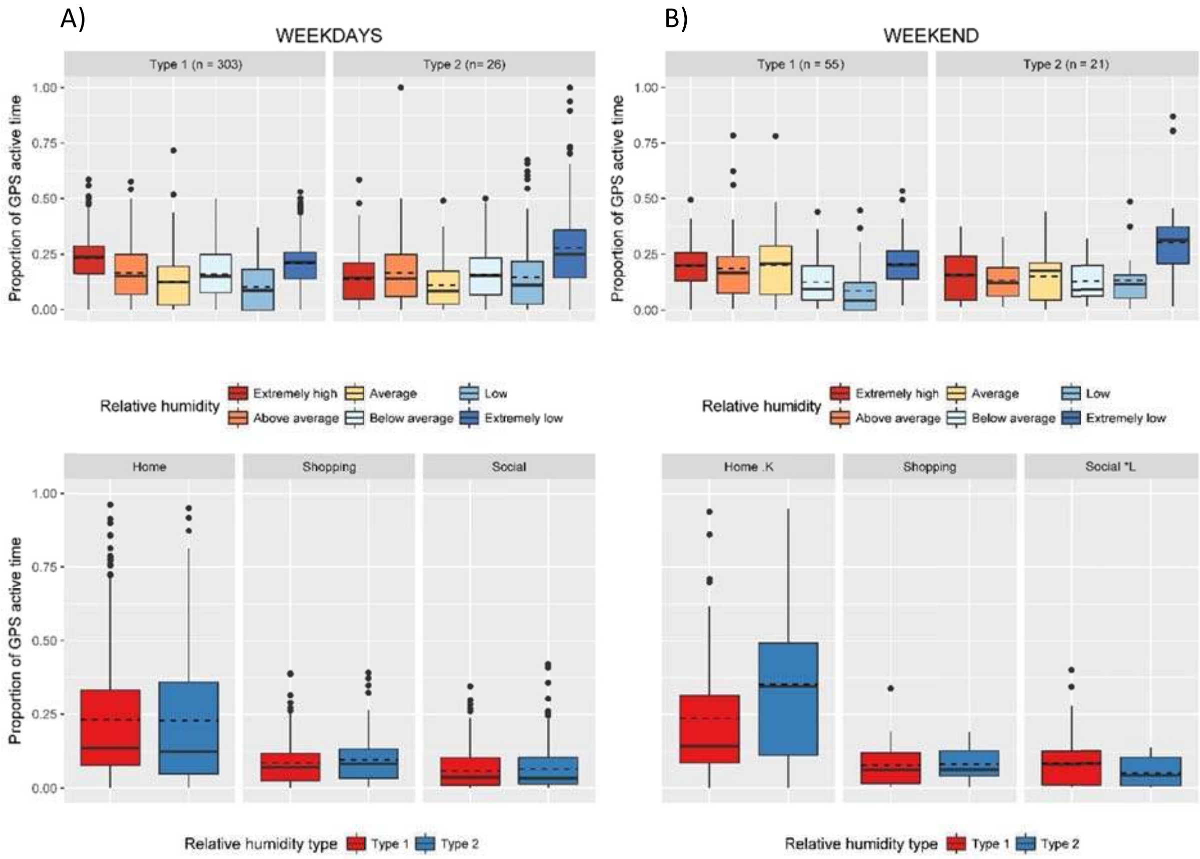

Figure 17 – Clusters for MCSA run on relative humidity and places on weekdays (A) and weekends (B). The two top panels describe relative humidity conditions within each cluster (Types) and respective proportions of GPS active time spent under relative humidity classes. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on places by relative humidity type described above.The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

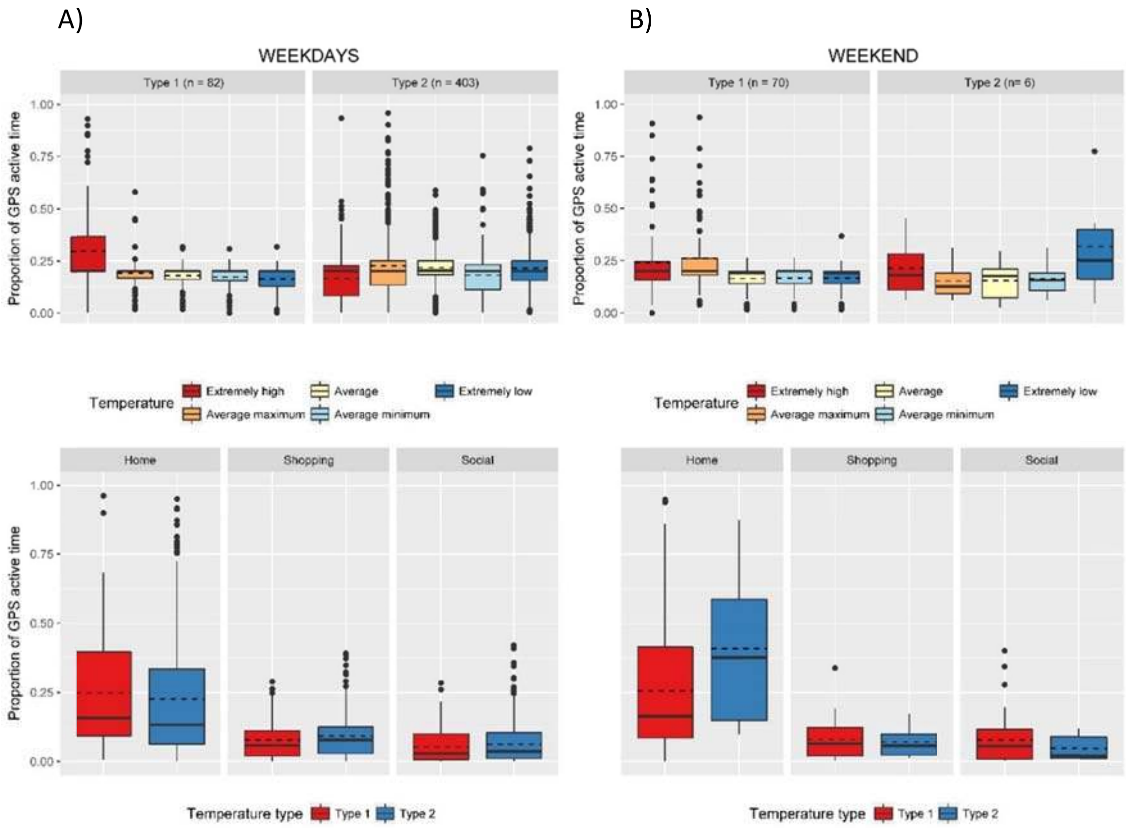

Figure 18 – Clusters for MCSA run on temperature and travel modes on weekdays (A) and weekends (B). The two top panels describe temperature conditions within each cluster (Types) and respective proportions of GPS active time spent under temperature classes. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on each travel mode by temperature type described above.The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.

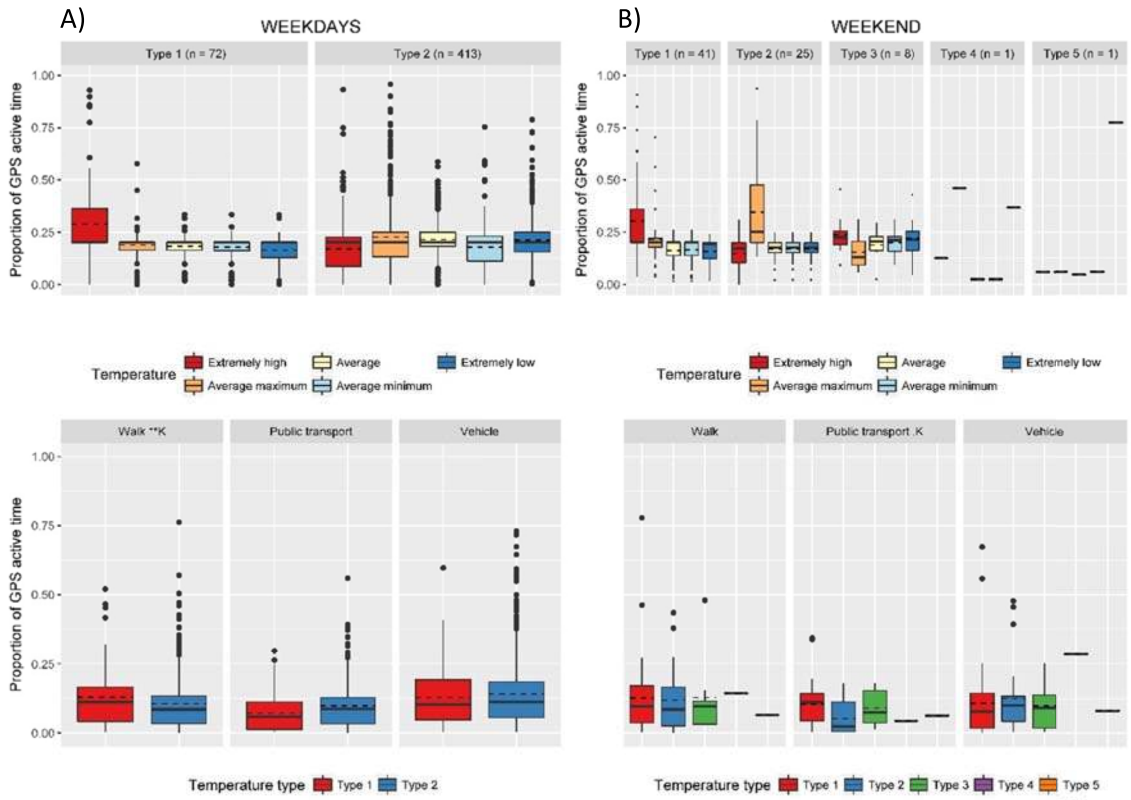

Figure 19 – Clusters for MCSA run on temperature and travel modes on weekdays (A) and weekends (B). The two top panels describe temperature conditions within each cluster (Types) and respective proportions of GPS active time spent under temperature classes. The two panels at the bottom show boxplots with the distribution of proportional GPS active time spent on places by temperature type described above. The dashed line on boxplots show the average and the continuous line the median. L reports significance from Levene's test and K from Kruskal-Wallis' test.