Investigating the Effect of Wind on Human Movement Behaviour using Multichannel Sequence Analysis

Vanessa S. Brum-Bastos [*1], Jed A. Long [†1] and Urška Demšar [‡1]

[1] School of Geography & Sustainable Development, University of St Andrews

December 15, 2017

Summary

The influence of the weather on human movement behaviour is of interest in urban planning and transportation research. Traditionally, this is investigated with travel surveys and questionnaires. In this paper we propose to use a combination of Context-Aware Movement Analysis (CAMA) and Multichannel Sequence Analysis (MSA) on real meteorological data and GPS tracking data. CAMA integrates movement data with spatio-temporally simultaneous meteorological information. MSA represents a person's movement as a multichannel sequence of states, where each channel describes either travel mode or the meteorological conditions. Similar movement patterns can then be identified by aligning sequences and finding their differences.

KEYWORDS: GPS tracking, environmental data, context aware movement analysis, multichannel sequence analysis, movement analytics.

1. Introduction

The influence of weather on travel mode is important in urban planning and transportation geography. Traditionally, these influences were explored through questionnaires and travel surveys (de Freitas 2003), but less so with real data. In this paper we propose to combine real meteorological data with GPS tracking data to investigate how weather affects the choice of travel mode. We do this through Context-Aware Movement Analysis (CAMA), which integrates movement geometry with the surrounding conditions that might be affecting movement (Dodge et al. 2013); and Multichannel Sequence Analysis (MSA, Gauthier et al. 2010), which represents movement as a sequence of characters, describing either the type of movement or the state of the environment throughout time.

Sequence alignments methods were developed for DNA research (Idury & Waterman 1995), in which DNA molecules are represented as a string of characters. The comparison of similarities between DNA strings allows the identification of protein sequences related to, for example, genetic diseases. The same principle can be applied to longitudinal data in social science for identifying groups of people with similar timelines (Billari 2001).

We apply CAMA and MSA to explore the effect of wind on human movement patterns. GPS tracking data are linked to wind strength and direction data and converted into multichannel sequences, which are clustered based on their similarity. These clusters are analysed and interpreted in terms of time spent using different travel modes.

2. Methods

2.1. Data and context-aware annotation of GPS tracks

[*] vdsbb@st-andrews.ac.uk

[†] jed.long@st-andrews.ac.uk

[‡] urska.demsar@st-andrews.ac.uk

We used tracking data recorded in a GPS survey of 91 participants in Dunfermline (Figure 1, Siła-Nowicka et al. 2016). Each participant was tracked for two consecutive weeks between 28/9/2013 and 10/1/2014 with a five seconds sampling frequency (Oshan et al. 2014). GPS points were pre-classified into travel modes: walking, public transport and vehicle use; and places. In this paper we focus on travel modes only.

Hourly contextual data on wind strength and direction came from the UK Met Office MIDAS meteorological stations system (UK Met Office 2012), which were interpolated in space using Thiessen Polygons and integrated to movement data via linear dynamic trajectory annotation (DTA-L, Brum Bastos et al. 2018) . DTA-L intersects in space the contextual layer immediately before and after in time for each GPS point and then uses a linear interpolation to find the contextual value at the time when the fix was collected.

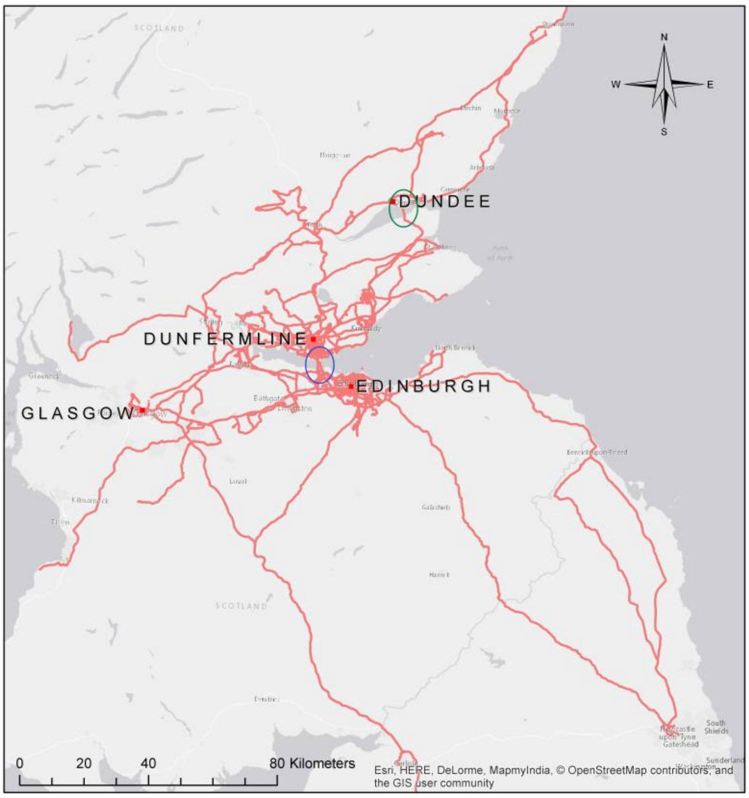

Figure 1 Area where the GPS data collection took place, trajectories are represented by the light red lines. The dark blue circle shows the Forth Bridges and the dark green circle shows the Tay bridges.

2.2. Representing movement with multichannel sequences

We created alphabets for sequence analysis with categories of wind variables (Table 2) and generated the sequences for each person and day of the week. We created a timeline with all the seconds within a day and filled the seconds for which a GPS fix was recorded with the correspondent modes and wind variables. We subsampled sequences to one-minute temporal resolution using the mode. Figure 3 illustrates part of a resultant multichannel sequence.

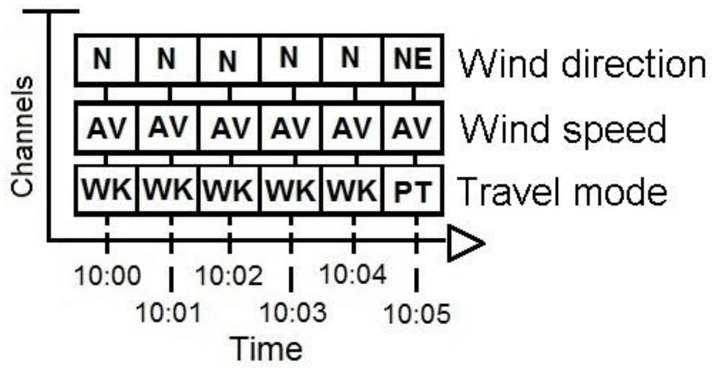

Figure 2 A Multichannel sequence for one participant over a five minutes period. Each channel relates to one of the wind variables or movement mode for each minute.

Table 2 Alphabets for wind variables and movement modes in multichannel sequences.

Wind Direction (compass degrees)			**Wind Speed (m/s)**		
Code	Description	Range	**Code**	Description	Range
N	North	> 337.5 - 22.5	**CM**	Calm	<= 3
NE	North East	>22.5 -67.5	**BR**	Breeze	>3 - 14
E	East	> 67.5 -112.5	**GA**	Gale	>14 - 24
SE	South East	>112.5 - 157.5	**ST**	Storm	>24
S	South	>157.5 – 202.5	**Travel mode**		
SW	South West	>202.5 – 247.5	**WK**	Walking	
W	West	> 247.5 – 292.5	**PT**	Using Public Transport (Bus, Train)	
NW	North West	>292.5 – 337.5	**V**	Vehicular use (Car)	

2.3. Multichannel Sequence Analysis (MSA)

Single channel sequence analysis has been used in movement analytics to identify groups of trajectories showing similar behaviour (De Groeve et al. 2016). We extend this approach by using multiple channels, where the similarity between two given multichannel sequences is based on the Optimal Match (OM) cost, which requires a cost matrix for each channel in a sequence. We used the transition rates calculated from the sequences for computing the cost matrices for mode, wind strength and speed, as shown in Equation 1.

$$F(i,j) = 1 - P(i,j) - P(j,i) \tag{1}$$

Here $F(i,j)$ is the cost and $P(i,j)$ is the transition rate from state i to j. The OM cost is calculated by aligning each channel in a sequence to all the other sequences in time and summing the costs of the of substitutions ($C_{s_i s_j}$), deletions and insertions (d) needed to turn one sequence into the other. The optimal alignment is the one with the smallest OM cost and can be computed using Equation 2.

$$F(i,j) = min\begin{cases} F(i-1,j-1) + C_{S_iS_j} \\ F(i-1,j) + \mathrm{d} \\ F(i,j-1) + \mathrm{d} \end{cases} \quad (2)$$

Here each line defines a possible OM score of two sub sequences, depending on which of the procedures, insertion, deletion or substitution, is less costly (Gauthier et al. 2010). $F(i-1,j-1)$ represents the optimal match 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 (Gauthier et al. 2010; Gabadinho et al. 2009). The substitution costs are given by a matrix C, where $C_{S_iS_j}$ is the cost for aligning S_i (the i^{th}character of I) against S_j, the j^{th} character of J. The OM cost between all multichannel sequences is represented by a K x K dissimilarity matrix, where K is the number of sequences.

2.4. Cluster analysis

We then used the similarity matrix in a clustering algorithm to find whether participants were showing similar movement behaviour under certain wind conditions. We used Ward's clustering (Murtagh & Legendre 2011), a hierarchical bottom-up algorithm that computes dissimilarities between two groups as the increase in the error sum of squares after merging those groups. We varied the number of clusters from 1 to the number of sequences and finally used the configuration with the highest Calinski-Harabaz Index (CHI, Everitt et al. 2011), which finds the best ratio between the within and between clusters dispersion.

To interpret the resulting clusters, we looked at the distribution of the proportion of time spent in different travel modes for the wind conditions associated with each cluster. We also 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 was enough evidence to support the existence of groups with different movement behaviour.

3. Results and discussion

There were no significant differences in the average time spent in different travel modes on weekdays under different wind conditions, on the weekend however we found significant differences in the average time expenditure in public transport and vehicles (Figure 3). The weekend wind Type 1 does not show a prevailing direction and the strength alternates between calm and gale for around 88% of the time; the weekend wind Type 2 comes predominantly from North-East, North-West and South-West, and alternates equally between calm, breeze and gale. There is a decrease in the average time spent on public transport under wind Type 2 on weekends and a concurrent increase on the average time expenditure in vehicles.

We speculate that this may be related to traffic restrictions at the Tay and Forth bridges (Figure 1) under high winds, which are more likely to come from NW, SW and NE in this location. We compared our results with the information from Traffic Scotland and found that there were at least ten occasions during our data collection during which the bridges were either closed or open only for cars because of high winds. Since the participants in our study were mostly commuters from Fife to Edinburgh or Dundee (Sila-Nowicka et al. 2016) and cross those bridges daily, it is possible that these restrictions are reflected in the patterns found through our CAMA+MSA approach. Further work, such as follow up interviews or travelling diaries, could help understanding the extent to which real behavioural patterns are reflected in our data set and ascertain the relationship between weather and movement behaviour.

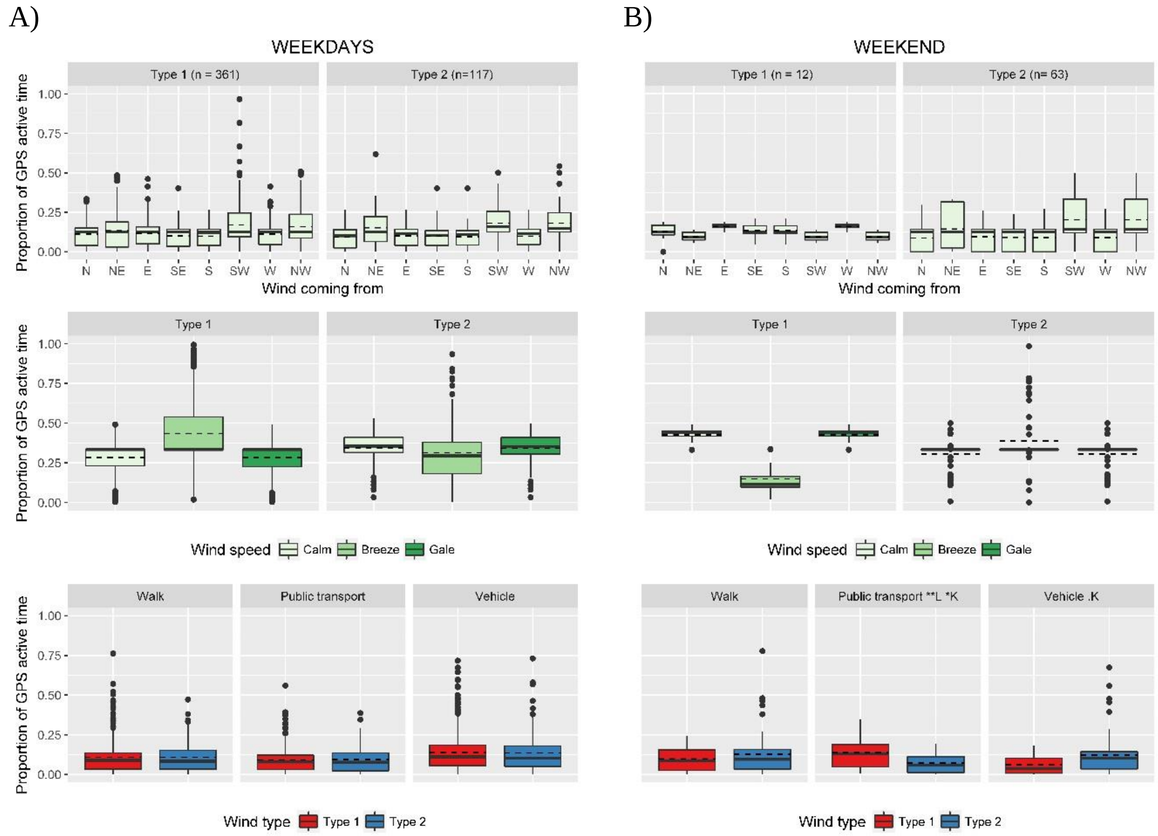

Figure 3 Clusters for MSA on wind speed, wind direction and travel modes on weekdays (A) and weekends (B). The four top panels describe the wind conditions within each cluster (wind types) and the respective proportions of GPS active time. The two panels at the bottom show the distribution of proportional GPS active time spent by wind type and travel mode. The dashed line is the average and the continuous line is the median. L reports significance (*** for α = 0.001, ** for α = 0.01, * for α = 0.5, . for α = 0.1) from Levene's test and K from Kruskal-Wallis' test.

4. Conclusions

We introduced a new data-driven approach that combines CAMA and MSA to identify the effects of weather on human mobility patterns. We ran this methodology on semantically annotated GPS tracking data with information on wind speed and direction to see if and how wind affects the choice of travel mode. This is work in progress and we are currently analysing the impact of further meteorological variables and their relevance for the choice of activity. The impact of meteorological conditions on activities has been initially explored through spatial interaction models (Sila-Nowicka & Fotheringham 2016), but MSA provide us with an alternative tool to perform such analysis.

5. Acknowledgements

The authors would like to acknowledge the SWB programme (Science Without Borders) for the financial support during the development of this study. The United Kingdom Meteorological Office (UKMO) provided the MIDAS - UK Hourly Weather Observation Data under the following licence: http://licences.ceda.ac.uk/image/data_access_condition/ukmo_agreement.pdf.

6. Biography

Vanessa Brum-Bastos is PhD candidate in GeoInformatics at the School of Geography and Sustainable Development at the University of St Andrews. She has a BSc in Geography from the São Paulo State University and a MSc in Remote Sensing from the National Institute for Space Research (INPE), Brazil.

Jed Long is Lecturer in GeoInformatics at the School of Geography and Sustainable Development at the University of St Andrews. He obtained his PhD and MSc in Geography from the University of Victoria. He also holds a BSc (Guelph, Canada) and an Adv. Dip. in GIS (COGS, Canada).

Urška Demšar is Lecturer in GeoInformatics at the School of Geography and Sustainable Development at the University of St Andrews. She has two degrees in Applied Mathematics from the University of Ljubljana, and a PhD in GeoInformatics from the Royal Institute of Technology (KTH), Stockholm, Sweden.

References

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), pp.1347–1370.

Billari, F.C., 2001. Sequence Analysis in Demographic Research. *Special Issue on Longitudinal Methodology Canadian Studies in Population*, 28(2), pp.439–458. Available at: http://www.canpopsoc.ca/CanPopSoc/assets/File/publications/journal/CSPv28n2p439.pdf [Accessed September 13, 2017].

Demšar, U. et al., 2015. Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*, 3(1). Available at: http://www.movementecologyjournal.com/content/3/1/5.

Dodge, S. et al., 2013. The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data. *Movement Ecology*, 1(1), p.3. Available at: http://www.movementecologyjournal.com/content/1/1/3 [Accessed March 31, 2014].

Gabadinho, A. et al., 2009. Mining Sequence Data in R with the TraMineR package: A user's guide. , p.144. Available at: https://cran.r-project.org/web/packages/TraMineR/TraMineR.pdf [Accessed September 14, 2017].

Gauthier, J.-A. et al., 2010. 1. Multichannel Sequence Analysis Applied to Social Science Data. *Sociological Methodology*, 40(1), pp.1–38. Available at: http://journals.sagepub.com/doi/10.1111/j.1467-9531.2010.01227.x [Accessed September 18, 2017].

De Groeve, J. et al., 2016. Extracting spatio-temporal patterns in animal trajectories: An ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, 7(3), pp.369–379.

Idury, R.M. & Waterman, M.S., 1995. A New Algorithm for DNA Sequence Assembly. *Journal of Computational Biology*, 2(2), pp.291–306. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7497130 [Accessed September 13, 2017].

Met Office, 2012. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). Available at: http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0 [Accessed September 18, 2017].

Oshan, T. et al., 2014. *GEOCROWD - Creating a Geospatial Knowledge World : Spatiotemporal patterns D1.9*, Available at: http://www.geocrowd.eu/wp-content/uploads/2015/01/D1.9_USTAN.pdf [Accessed September 8, 2017].

Siła-Nowicka, K. et al., 2016. Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5), pp.881–906. Available at: http://www.tandfonline.com/doi/abs/10.1080/13658816.2015.1100731.

Sila-Nowicka, K. & Fotheringham, A.S., 2016. A route map to calibrate spatial interaction models from GPS movement data. In *GIScience 2016 Workshop on Analysis of Movement Data*. Montreal, Canada.