Deliverable 4.7

Initial modelling of non-motorised transport

Grant Agreement number: 688082

Project acronym: SETA

Project title: An open, sustainable, ubiquitous data and service ecosystem for efficient, effective, safe, resilient mobility in metropolitan areas

Funding Scheme: H2020-ICT-2015

Authors

Ben Heller B.Heller@SHU.AC.UK

Internal Reviewer

Lyudmila Mihaylova l.s.mihaylova@sheffield.ac.uk

Matthew Hawes m.hawes@sheffield.ac.uk

State: Final version

Distribution: Public

Deliverable History

Date	Author	Changes
31/03/17	BH	First version
27/04/17	BH	Final version

Contents

1. Summary

Non-motorised mobility (NMM) is important for reducing urban traffic congestion, decreasing traffic emissions and improving public health. Recent developments in smartphone and sensor technology provide large-scale measurement of activity type and location, potentially allowing observation of the NMM behaviour of an entire city.

As part of the SETA project, we will use these data to develop models of NMM for decision makers. The models will:

1. Explore current non-motorised activity levels and behaviour.
2. Understand the effects of different determinants of NMM behaviour.
3. Use these determinants to produce *what-if* predictive models to:
    a. Predict future infrastructure demands due to changes in NMM.
    b. Predict the impact of new infrastructure schemes on NMM levels.
    c. Understand how NMM will integrate with the wider transport network and
    d. Help predict interactions between flows of cyclists, pedestrians, cars and other road users.
4. Classify journeys into different purposes (work, shopping, leisure, etc.).
5. Explore the interaction of different traffic modes on mode and route choice.
6. Quantify the effect of cycle parking and docking station availability for public bike rental on mode choice.
7. Explore how crime statistics affect non-motorised modes.
8. Determine the relationship between killed or seriously injured statistics for cyclists and pedestrians to vehicle speed.
9. Investigate the effect of escalators and moving walkways on cyclist and pedestrian behaviour.

The *social-ecological model* considers that policy, the social and physical environment and individual factors all influence behaviour. SETA will attempt to model measurable behavioural outputs (route and mode choice around NMM), we will use geographical and other information sources to determine physical environmental factors, but social and individual factors are outside the scope of SETA data collection and so must be inferred. We will use machine-learning techniques to identify clusters of NMM behaviour (representing patterns of response to observable determinants) or *personalities*. These personalities will be used within agent-based simulations of NMM journeys.

The stages of developing a model include identifying discrete journeys, understanding the purpose of that journey (commuting, shopping, leisure, etc) and then using observed determinants from SETA and other data sources to generate both descriptive and predictive models. These models will both explain and predict NMM behaviour.

As SETA data is not yet available, code and analysis techniques have been developed and piloted on preliminary and similar databases. Early results are described, and some procedural issues have been identified that will be addressed before roll-out of the full SETA data collection.

2. Glossary of Terms

Annex I	Otherwise known as the DoW
CA	Consortium Agreement
CVD	Cardiovascular disease
DoW	Description of Work
GA	Grant Agreement
EC	European Commission
PA	physical activity
IMU	inertial measurement unit
GIS	Geographical information system
KSI	Killed or Seriously Injured
NMM	Non-motorised mobility
TPB	Theory of planned behaviour
SOM	Self organising map
WP	Work Package

3. Introduction

3.1. Background – SETA

Urban mobility is of growing concern to citizens. Nine out of ten EU citizens believe that the traffic situation in their area should be improved. The choices that people make in the way they travel will affect not only future urban development but also the economic well-being of citizens and companies. It will also be essential for the success of the EU's overall strategy to combat climate change, achieve the 20-20-20 objective and to promote cohesion. (EC, 2009)

SETA is a project funded under the Horizon 2020 scheme (ICT-16-2015 - Big Data Research). It aims to use novel high-volume, high-velocity, multi-dimensional, heterogeneous data streams from existing traffic sensors, cameras, mobile phones and other sensors to understand, optimise and manage urban mobility and make it more efficient, sustainable and resilient. SETA's solution will be implemented and evaluated by citizens, business and decision makers in three cities across Europe (Santander, Turin and Birmingham).

As well as improving the granularity, accuracy and timeliness of conventional traffic modelling, this new approach will obtain data about non-motorised mobility (NMM - cycling, walking, running, etc.) and mixed-modes of transport (park and walk, park and cycle, train and cycle, etc.). In this way, a totally integrated model of urban transportation will be created that will foster sustainable types of mobility, especially walking, cycling and public transport.

3.2. Non-motorised transport

The green paper "Towards a new culture for urban mobility" (EC, 2007) identifies several issues affecting urban transport:

1. **Congestion**: 1% of the EU's GDP is lost to the European economy because of congestion.
2. **Emissions**: urban traffic is responsible for 40% of CO2 emissions and 70% of emissions of other pollutants. Traffic noise can lead to poor quality of life in residents near busy roads.
3. **Safety**: In addition, and a potential barrier for wider acceptance of non-motorised transport, one in three fatal accidents happen in urban areas, and it is the most vulnerable people, namely pedestrians and cyclists, who are the main victims.

All of these issues are addressable through a wider uptake of non-motorised transport, together with appropriate infrastructure changes. In addition to the impact on urban transport, active travel (walking running or cycling for commuting or other utilitarian purposes) has individual health benefits and is one of the most acceptable ways to meet government recommended levels of physical activity (PA), which is associated with a 30% risk reduction in all-cause mortality (Department of Health, 2011). Studies suggest that obesity is positively associated with motorized vehicle travel or negatively associated with active travel (or both) across a broad cross-section of nations (King and Jacobson 2017). A recently released study of 264 277 UK participants with median 5 year follow-up (Celis-Morales, 2017) has shown that cycling to work is linked to a substantial decrease in the risk of developing and dying from cancer or heart disease and all-cause mortality. People who walked to work, covering more than six miles a week had a lower risk of CVD incidence; they also had a lower risk of CVD mortality, but walking had no statistically significant impact on all-cause mortality or cancer outcomes. A study in Barcelona of the *Bicing* public bicycle sharing initiative found an annual increase in mortality for Barcelona residents of 0.03 deaths from road traffic accidents and 0.13 deaths from air pollution, but a reduction of 12.46 deaths annually due to increased physical activity, giving a benefit ratio of 77 (Rojas-Rueda, 2011).

An editorial in the British Medical Journal referring to the Celis-Morales study states:

> *The findings from this study are a clear call for political action on active commuting, which has the potential to improve public health by preventing common (and costly) non-communicable diseases. A shift from car to more active modes of travel will also decrease traffic in congested city centres and help reduce air pollution, with further benefits for health*. (Andersen, 2017)

Accurate, long-term data are needed in order to accurately determine the health and economic consequences of active travel, and to examine the effectiveness of any interventions. However, Saunders *et al* (2013) identify limitations of studies of transportation choice to date:

1. Lacking long-term data describing physical activity and travel behaviours.
2. Having limited generalizability due to geographically limited scope.
3. Being open to potential bias by relying on self-reported data.

Recently developed measurement devices combining GPS, global information systems (GIS) and inertial measurement unit (IMU) sensors allow researchers to capture objectively measured data about a participant's location, surrounding environment, and physical movements. However, these sensors are limited with practical issues including limited battery life and unreliable satellite signals (Oliver *et al.* 2010). Modern smartphones have effectively solved these problems by combining long-battery-life, IMU sensors and redundancy of location detection[1], allowing large-scale, long-term, geographically distributed studies, with the additional benefit of near-real-time data upload. SETA will utilise these and other technologies (e.g. automated analysis of video footage) to enable large-scale, timely, long-term and objective measurement of active travel.

3.3. Urban mobility prediction models

SETA Deliverable 4.1 -*Exploring prediction perspectives* (SETA 2016) reports on state-of-the art modelling techniques as applied to motorised traffic prediction. This report proposes three applications of modelling of traffic flow:

1. Dynamic traffic management, i.e. re-routing traffic through less congested paths, or regulating traffic light control in order to better accommodate the traffic flow.
2. Context-aware navigation apps to estimate and compare traffic conditions along alternative paths towards the selected destination.
3. Use of the collected data by decision makers for purposes ranging from transport (service) planning and design, policy evaluation and monitoring, and uses beyond the mobility domain itself, e.g. asset management, city planning, etc.

Of these three applications, only the third will be considered for this deliverable. Congestion of non-motorised vehicles is generally less critical than for cars due lower uptake and the bicycle's lower space requirements, for example a passenger-car-unit of 0.2 is used for bicycles, vs 1.0 for a passenger car (Smith and Blewitt, 2010). Any congestion that does occur is less-easily managed by traffic-light

[1] Through fusion of multiple satellite location, local Wi-Fi and cellular base-station locations.

control and instead requires long-term infrastructure intervention. Similarly, route-choice (and hence navigation apps) are likely to be less affected by the density of other cycles and more by motor-vehicle densities on the route. The same arguments apply to pedestrian travel. This situation is the case for the three SETA cities; other urban areas with higher active travel densities may also require applications 1 and 2.

3.4. Urban non-motorised mobility prediction models for decision makers

Relevant applications of the SETA NMM model include:

1. Exploring current non-motorised activity levels and behaviour (e.g. for policy evaluation, investigating the effect of any interventions, compensating for seasonal and environmental influences).
2. Understanding the effects of different determinants of NMM behaviour and hence developing and calibrating behavioural models to allow what-if prediction
3. Using the what-if predictive models to:
    a. Predict future infrastructure demands due to projected increases in NMM.
    b. Conversely, investigate the potential impact of new infrastructure schemes on NMM. This is particularly important when the potential costs and benefits of multiple competing interventions are to be compared.
    c. Understand how proposed infrastructure schemes will integrate with the wider transport network and
    d. Help predict interactions between flows of cyclists, pedestrians, cars and other road users.

3.5. Cities' specific requirements from non-motorised mobility prediction models

Specific information requested by decision makers within the SETA project include:

1. Classifying active travel journeys into different purposes (work, shopping, leisure, etc.).
2. The effect of volume of all modes on cyclist or pedestrian route selection (so if high volume of people walking do they divert? Or if there is traffic, how does this affect their choice? Including the effect of vehicular speed and size on sustainable travel modes.
3. How cycle parking availability and the availability of docking stations for public bike rental may affect mode choice.
4. How crime statistics affect non-motorised modes.
5. The relationship between killed or seriously injured (KSI) statistics for cyclists and pedestrians to vehicle speed.
6. Investigate the effect of escalators and moving walkways (in Santander) on cyclist and pedestrian behaviour.

4. Determinants of non-motorised mobility choices

4.1. The Social-ecological model

Physical activity behaviour and the factors influencing it are extremely complex. Models are used to generalise, simplify and help clarify the factors (determinants) that act as facilitators or barriers to participation in physical activity. Models help to identify determinants of physical activity participation in specific populations therefore enabling the design of effective interventions (strategies, programs or policies that increase physical activity participation).

The social-ecological model is a comprehensive approach to designing, implementing and evaluating interventions, including those targeting physical activity behaviour. The term ecology comes from biology, referring to the interrelationships between organisms and their environments The social-ecological model considers that policy, the social and physical environment and individual factors all influence behaviour and so change is more likely to be successful when these multiple levels of influence are addressed simultaneously. This is illustrated graphically in Bauman's socio-ecological model of cycling behaviour (Bauman et al., 2008) shown in Figure 1.

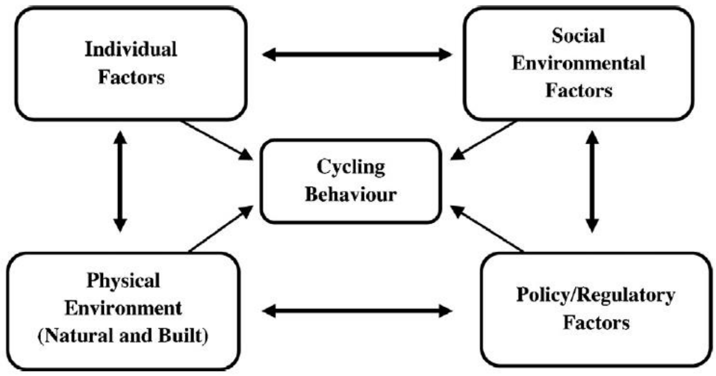


Figure 1- Social ecological model of cycling behaviour from Bauman et al. (2008)

Rowe et al. (2013) present some of the socio-ecological factors that may influence cycling for transport behaviour, these are listed in Table 1.

Socio-ecological construct	Relevant factors
Individual factors	Age, gender, confidence, knowledge, ability
Social environmental factors	Social norms for transport-related cycling, knowing others who travel by bicycle, driver culture, driver behaviour
Physical environmental factors	Safety, road conditions / degree of separation from cars, weather, terrain, end of trip facilities

Policy factors	Road rules, funding for path and road infrastructure, urban planning policies, active transport policies

Table 1 factors affecting cycling active travel from Rowe et al. 2013

One of the key aspects of the socio-ecological model is that the relationships are dynamic, they can change with time and interact. For example, changes made to the environment can influence individual behaviour (such as increasing the number of people cycling), which can alter the social environment, which can influence government, which can modify policy which in-turn leads to environmental changes, etc. These interactions are illustrated in Figure 1. The critical insight of the social ecological perspective of human behaviour is the need to maximise the 'person-environment fit' in order to ensure the best opportunity for behaviour change.

As the policy construct is not responsive in the short-term, and is often outside the control of decision makers, it will not be considered further in this model. The remaining socio-ecological constructs are explored in more depth below.

4.2. Individual factors

Using Michie's framework (Michie *et al*., 2011), the individual construct can be further modelled as factors representing:

1. **Capability:** both our psychological and physical capacity to enact the behaviour, including the necessary knowledge, fitness and skills.
2. **Opportunity**: an individual physical and social environment that permits the behaviour to occur. Owning a bike, having a short commute and changing facilities in one's work environment presents the opportunity for cycle commuting to take place.
3. **Motivation:** reflective and automatic mechanisms that activate or inhibit behaviour, they can be intrinsic motivators - doing an activity because it is inherently satisfying (e.g. the enjoyment of cycling), or extrinsic motivators - doing an activity for instrumental reasons (e.g. for weight loss or to save money). Many psychological models of motivation have been applied to physical activity behaviour. One commonly applied model that seeks to explain the role of social aspects in an individual's behaviour is Ajzen's theory of planned behaviour (TPB, in Rivis and Sheeren, 2003). This postulates that intentions to perform a behaviour are affected by three factors i) attitudes towards the behaviour, ii) perceived behavioural control (the perception of the ability to perform that behaviour) and iii) subjective norm (the belief that others desire the behaviour). Extending the TBP by adding the *descriptive norm*, a measure of the perception of what other people are doing, improved prediction of behavioural intent (Rivis and Sheeren, 2003).

Influences at the individual level can be on the *capability* dimension including cycle training and maintenance (e.g. bike doctors); the *opportunity* dimension such as provision of bicycles (e.g. the Big Birmingham Bike Giveaway – BBC 2015) and can target *motivation* such as provision of rewards for active travel, increasing charges for car parking, etc.

4.3. Social-environmental factors

Of the many cues that influence behaviour, at any point in time, none is more common than the actions of others. (Bandura, 1986, in Ball, 2006)

A review of physical activity correlates in adults found that four social factors (social support from friends/peers; from spouse/family; physician influence; and social isolation) were consistently important in predicting physical activity (Trost *et al.* 2002). This supports the importance of strengthening social influences in physical activity promotion – encouraging activity with family/friends, or group participation, and especially fostering cultures in which physical activity is normalised, that is commonly undertaken/observed and socially desirable.

An investigation of the influence of the social environment on walking for transport was undertaken by Clark and Scott (2013). They investigated 4 components of the social environment: companionship (walking with other people), encouragement (when family, friends, or other acquaintances promote walking), role models (people who walk and whose own participation encourages others to become involved) and social cohesion (the extent to which a neighbourhood is socially interconnected). Of these social variables, role-models (specifically seeing friends participating in walking for transport) and social cohesion ("People in my neighbourhood can be trusted") were both significantly associated with walking for transport.

One interesting consideration of social-environmental factors is that the relative risk of injury falls with increased cycling or walking; the evidence seems to suggest a square root relationship or a 'safety in numbers'- effect. It is not clear if this relationship is causal, or what the mechanisms might be, (e.g. presence of more cyclists might make drivers more cautious, conversely the provision of safer cycling routes might encourage more cycling).

The development of social networking services has allowed social influence to extend beyond immediate neighbourhood and family / friends. This allows creation of communities of common mutual interest, providing a strong social norm (subjective and descriptive) for performance of the behaviour in question, even if this is not common in the user's real-world community. Strava (a "social network for athletes") is an example of the use of social networks to communicate physical activity and receive encouragement.

4.4. Physical-environmental factors

The physical environment refers to the design of the urban landscape. Brownson *et al.* (2009) suggest that different aspects of physical activity (e.g., leisure, transportation, household) are affected by different environmental attributes. Leisure physical activity may be most affected by access to, and characteristics of, public and private recreation facilities. Transportation physical activity may be most related to the proximity and directness of routes from home to destinations as well as characteristics of the walking and cycling infrastructure. They suggest that to understand the influences of the built environment on physical activity, a wide range of environmental measures are needed. They detail the four most common measures of the physical environment that influence walking: population density, land-use mix, street connectivity, and sidewalk (pavement) availability. A more recent comprehensive review by Sallis *et al.* summarised 418 studies investigating five physical activity settings (parks/open space/trails, urban design, transportation, schools, and workplaces/buildings), with multiple, evidence-based, activity-friendly features identified for each setting. Six potential outcomes were considered (physical health, mental health, social benefits, safety/injury prevention, environmental sustainability, and economic). Environmental features with the strongest evidence of multiple co-benefits were found to be park proximity, mixed land use, trees/greenery, accessibility and street connectivity, building design, and workplace physical activity policies/programs.

Ewing and Cervero (2010) categorise measures of the built environment into *D variables:*

Density - population, dwelling units, employment, building floor area, or other variables measured per unit of area.

Diversity measures reflect the number of different land uses in a given area. For entropy measures of diversity, low values indicate single-use environments and higher values more varied land uses.

Design includes street network characteristics (connectivity) within an area. Measures include average block size, proportion of four-way intersections, and number of intersections per unit area. Design is also measured as sidewalk coverage; average building setback; average street width; or numbers of pedestrian crossings, street trees, or other physical variables that differentiate pedestrian-oriented environments from car-oriented ones.

Destination accessibility measures ease of access to end points. It may be regional or local. In some studies, regional accessibility is distance to the central business district. In others, it is the number of jobs or other attractions reachable within a given travel time. Local accessibility is distance from home to the closest store. Distance to transit is usually measured as an average of the shortest street routes from the residences or workplaces in an area to the nearest rail station or bus stop.

Two additional *D* variables sometimes reported are *Demand management*, including parking supply and cost. While not part of the environment, *demographics* is controlled as a confounding influence in travel studies.

Attempts have been made to combine physical environmental measures into single indexes, for example *walkability* indexes, combining multiple components of the physical environment into a single variable that is then used to assess impact on walking. Frank *et al*. (2010) developed one such index combining net residential density, intersection density, land-use mix, and retail floor area ratio; the index was found to be significantly related to walking prevalence.

Although less researched, cycle-ability is similarly assessed through measures of connectivity, exposure to traffic, cycle parking, etc. For example, the UK's Transport Research Laboratory produce street audit software that combines evaluation of both cycling and walking environments (https://trlsoftware.co.uk/products/street_auditing).

Traditional transport modelling methods may have limited application in the modelling of active travel due to their lack of granularity, which ignores the small-scale features affecting the travel decisions of pedestrians and cyclists (Cooper, 2017). Spatial network analysis (which works on the un-simplified network level) has been used as a technique to predict cycling flows, accidents involving non-motorists and accessibility to public transport. Cooper suggests that the use of angular betweenness (a measure of the *straightest* rather than the *shortest* path) which is often used for prediction of vehicle flows, is less relevant for cyclist. Instead, Cooper proposes using distance, slope, vehicle traffic and angular distance. Cooper argues that this betweenness measure alone can replace the 4 stages of the traditional transport model (trip generation, trip distribution, mode choice and route choice). Cooper successfully adapted Broach *et al.'*s (2012) model of cyclist route choice in Portland, USA to fit a UK city (Cardiff) with different street layouts and information availability.

Despite the success of walkability and cycle-ability models in predicting aspects of active travel, it is necessary to recognise that social-environmental factors are only one of the determinants influencing mode choice. The importance of the social-ecological model (section 4.1) is illustrated by studies that show no effect of walkability measures on objectively measured physical activity for certain populations (Van Cauwenberg *et al.*, 2011, Carter *et al.*, 2017). In addition, there is some evidence that, at least in the short-term, new walking or cycling infrastructure may merely displace activity from other routes (Goodman, *et al.*, 2014).

4.5. Implications for SETA

It has been demonstrated that behaviour around active travel is complex and multi-dimensional, dependent on a multiplicity of individual and environmental factors. Due to the necessity of minimising the burden of data collection on SETA users, many of these determinants (particularly individual and social) will not be accessible within the SETA data ecosystem. However, compared to previous studies, SETA provides the opportunity of collecting long-term transport mode and route choice information across a large cross-section of users in three urban areas, each providing a variety of social and physical environmental situations. In this way, it constitutes a *natural experiment* - events, policies, or interventions not designed for research purposes but that may nevertheless provide valuable research opportunities (Craig *et al.*, 2011). In particular, *clustering* techniques, will be used to identify *personalities* (patterns of behaviour in the data), rather than assuming *a-priori* determinants. The personalities will represent mediators between observable determinants and behavioural responses and will be used for agent-based simulation of non-motorised transport.

5. Workflow

5.1. Overall workflow

The overall structure of the NMM modelling within SETA is shown in Figure 2. It shows all the data sources around *journey contextualisation* – understanding the journey in its social-ecological setting.

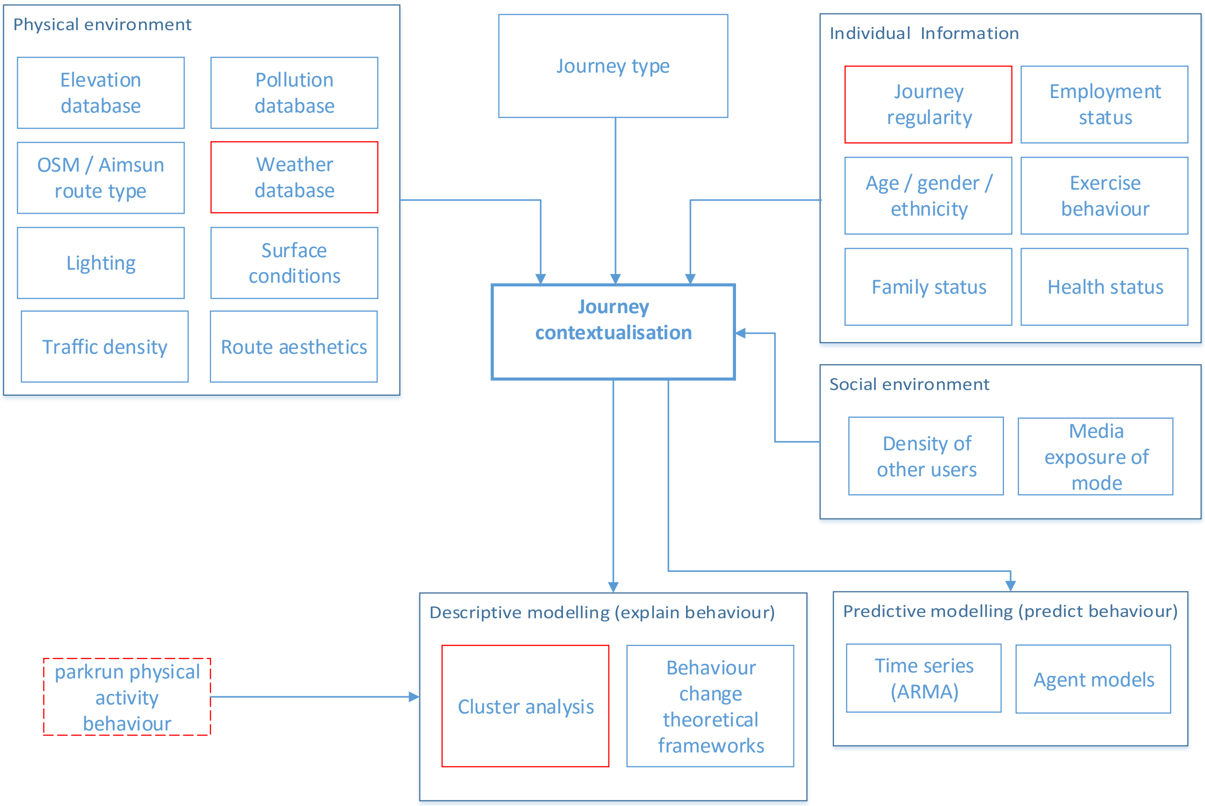


Figure 2- behavioural modelling – current work is shown in red

Journey type represents the classification of a journey's purpose (commuting, shopping, other utilitarian, leisure). This has been shown to be an important factor in mediating the influence of other determinants on the mode and route choice (Broach *et al*., 2012). The process of identifying journey type is illustrated in Figure 3. Finally, the process of identifying individual journeys is illustrated in Figure 4.

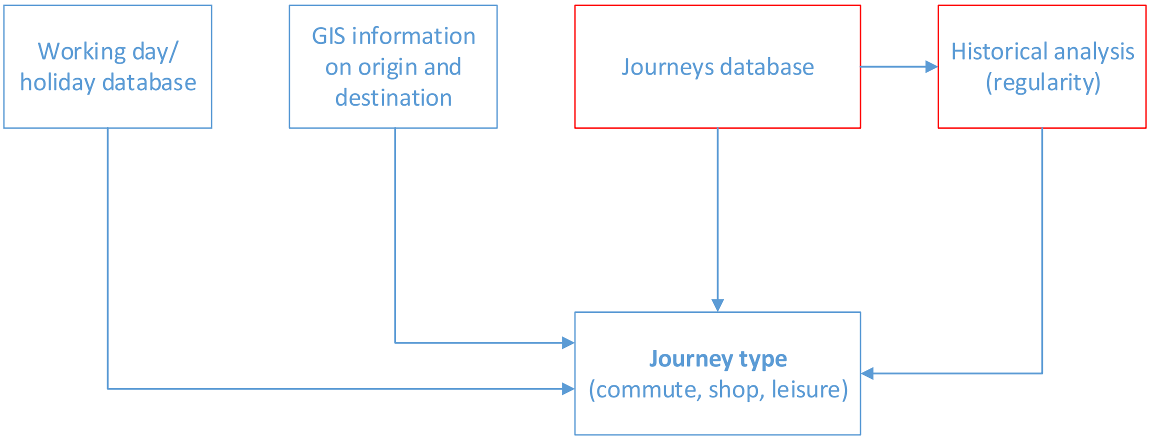


Figure 3: journey type identification – current work is shown in red

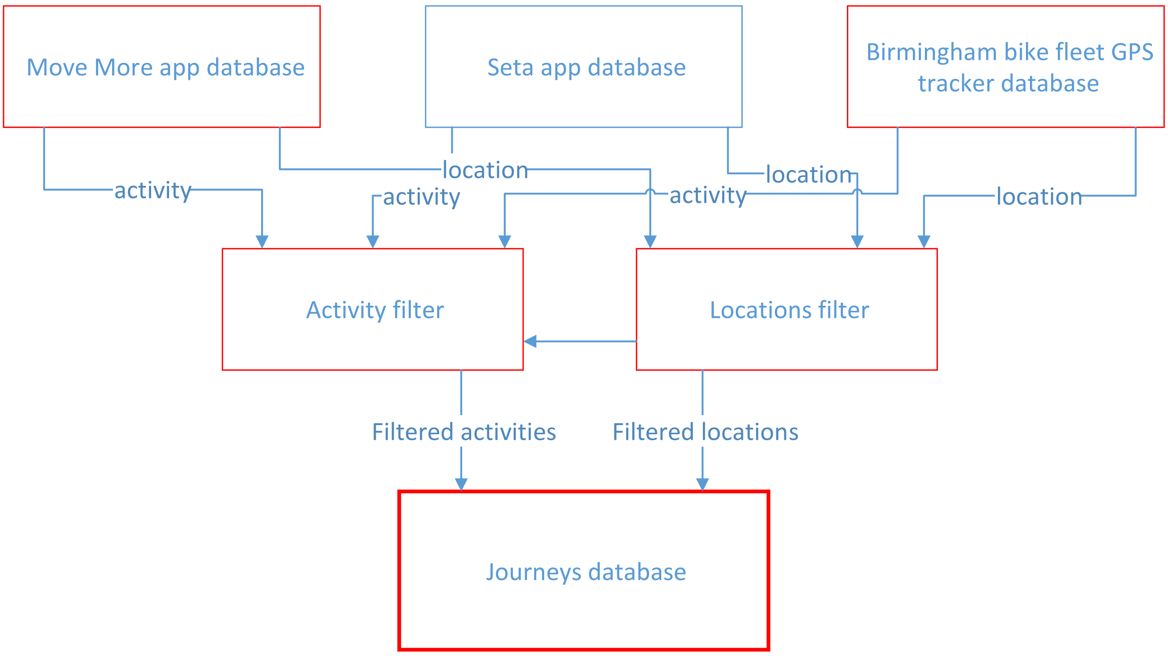


Figure 4: journey identification - current work is shown in red

5.2. Initial workflow

This is a preliminary report, covering initial modelling of NMM. Due to the relatively early stage of the SETA project, detailed SETA mobility data is not yet available. However, in-order to pilot the processes required to perform the full modelling (to be reported in D4.8 - Final modelling of non-vehicular transport – month 33), code and analysis techniques have been developed and tested on preliminary and similar databases. These are shown in red in the flowcharts and are described in the preliminary results section (7).

6. Datasources

In anticipation of the availability of SETA mobility data, other databases have been used to provide similar data on mobility and physical activity at large scale.

6.1. Move More data

Move More is an initiative in Sheffield UK, designed to motivate citizens of all ages to be more active. The Move More Month took place in July 2016, prior to the Rio Olympics and consisted of a publicity drive, the provision of many free exercise opportunities, a workplace challenge and provision of a free app – the Move More Sheffield app (Workout-UK, 2016). The app was designed to run on Android (690 users) or iOS (1282 users) operating systems and record all activity without user interaction. Activity type, time and location are written to a database at regular intervals from every phone.

The app was developed by the OAK computing group at the University of Sheffield (partners in SETA), who also manage the database server. Due to its sensitive nature, Move More data has restricted availability, Sheffield Hallam University have been granted research ethics approval to analyse the data for research purposes.

Although the Move More Month finished at the end of July 2016, several users have continued to use the app, and another 190 users registered in the month of March for a subsidiary competition between the two Sheffield universities, thus the data include some information on seasonal factors.

Figure 5 Example of Move More NMM data – heat map of running data around Sheffield, July 2016.

6.2. Big Birmingham Bikes data

The Birmingham Cycle Revolution is a programme aimed at increasing cycle use in Birmingham, with an investment of around £24 million on infrastructure to make Birmingham a cycle-friendly city, and to increase cycle use to 5% of journeys in 10 years' time and 10% of journeys in 20 years' time. One component of the Cycle Revolution is a scheme called the Birmingham Bike Giveaway which has given away 4,000 bicycles to residents of the city in economically deprived areas. The majority of the bikes have automatic GPS bike trackers installed, which return location, date and time data whenever there is a change in speed, direction or the distance travelled exceeds 1km. Unfortunately, many of the trackers ceased functioning due to incorrect fitting or inadvertent damage. However, there are 185 bikes returning GPS data as of March 2017. Due to it's sensitive nature the Birmingham bike GPS data has restricted access. As SETA partners, Sheffield Hallam University has access to these GPS data.

6.3. parkrun data

parkrun is a UK-based charity that organises free weekly community events over a five kilometre course which participants can run or walk every Saturday morning, in a local park or other public space. parkrun HQ provides the strategic and practical management nationally and internationally, whilst events are organised locally by volunteer teams. When registering on the parkrun website, individuals are allocated a number and barcode which they bring to the event. At the end of the run, barcodes are scanned so that individuals' times are recorded and the results posted on the website and sent to individuals by email or text.

There are over 1100 weekly events worldwide, with approximately 150,000 weekly runners and over 10,000 volunteers. As of April 2017, over 23 million runs have been recorded. Whilst data for each even are public, individual data is sensitive and hence restricted. Sheffield Hallam University collaborates with parkrun to support research into exercise, wellbeing and social cohesion, and has been granted access to the parkrun historic database of over 2 million registrants. This provides information on age, current activity levels, postcode as well as weekly attendance and performance data, so is a valuable resource to conduct large-scale research into physical activity behaviour.

6.4. Weather data

Historical UK weather observation data is available to academic institutions from the Met Office (the UK's National Meteorological Service). This is used in our NMM models to understand the effect of weather (specifically temperature, rain and wind) on travel and physical activity decisions. The data are available freely to researchers and are described here http://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c.

6.5. Route type data

Open Street Map (OSM). OSM is an open access, crowd sourced mapping product with global scope; while coverage is currently patchy, both coverage and data quality are improving future, and OSM currently provides the best coverage of UK cycle paths (Lovelace, 2015).

See http://wiki.openstreetmap.org/wiki/Cycle_routes

6.6. Elevation data

Elevation data are taken from Ordnance Survey Terrain 50 (freely available).

(https://www.ordnancesurvey.co.uk/opendatadownload/products.html#TERR50)

7. Preliminary Results

The majority of work at this stage has focussed on understanding Move More journeys, as these data are in identical format (activity, location and time) to the anticipated SETA mobility data.

7.1. Tools

A variety of tools have been utilised to analyse and visualise the data. Data storage is predominantly in SQL format on a server. Microsoft Power BI has been used to quickly amalgamate datasources, with some processing in R. Daxstudio is used to output data to Matlab where further data processing takes place (such as clustering). Visualisation is currently through Power BI (Power View), Matlab or other software such as Excel as appropriate. Custom output visualisations will be produced in Workpackage 5 once the tools are further developed.

For modelling of cycling and walking flows, the open-source sDNA tool will be used (Cooper & Chiaradia 2015).

7.2. Journey filtering

A fundamental issue is developing a set of rules for deciding when to split journeys that include stationary periods. The GPS bicycle data from Birmingham was analysed for a range of different time thresholds and the effect on the resulting number-of-journeys was considered.

It can be seen from Figure 6 that (going from longer to shorter), there is a change in the rate at which the journeys increase at 18 minutes, and a second change at 10 minutes. The threshold is thus taken as the mean of these values at 14 minutes. This crude figure is suitable for preliminary analysis, but a more sophisticated analysis will be performed when more data are available to also include factors of journey purpose and journey length.

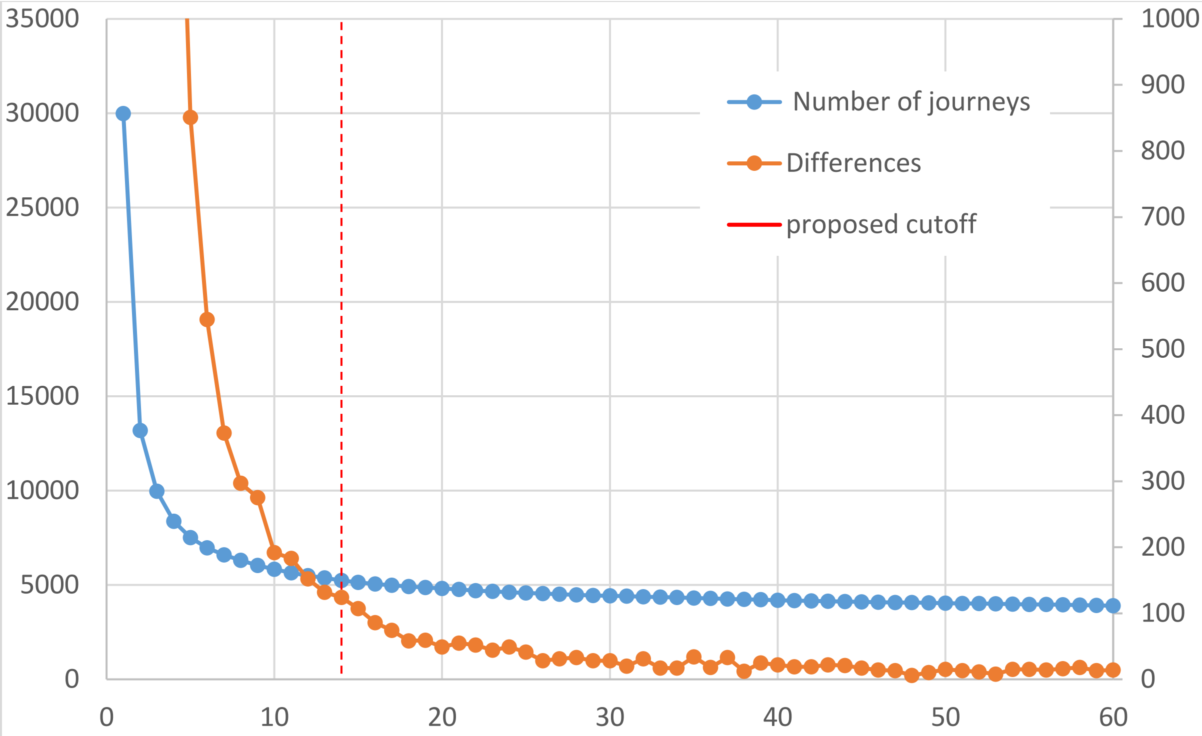


Figure 6: cut-off time (minutes) vs number of journeys. Differences is the difference in journey numbers between subsequent cut-off times.

7.3. Journey clustering

In order to be able to classify journey types (as shown in Figure 3**Error! Reference source not found.**) it is important to establish the regularity of the journey. This allows determination of both place of residence and place of work, along with other common destinations. GPS location data typically have error of 10s of metres, especially within crowded urban environments and at the start of journeys before a satelite fix is established, so a clustering technique is needed to group similar origin and destination points. A hierarchical clustering approach is used for this (Murtagh 1985).

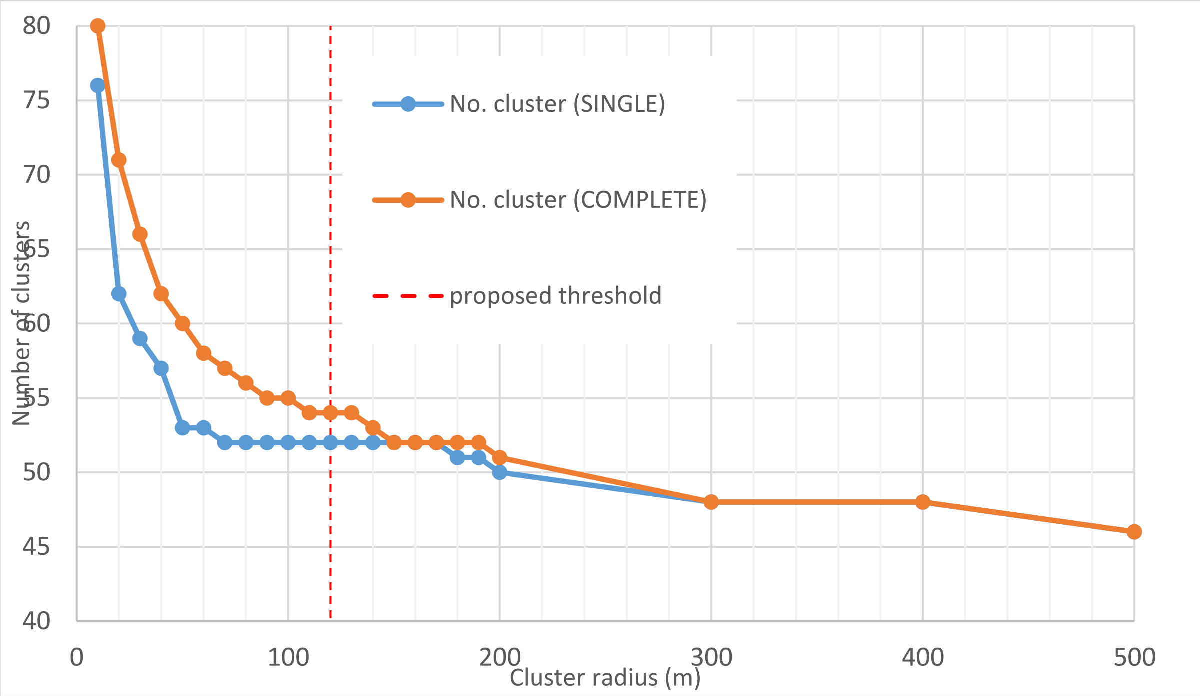


Figure 7: Number of clusters vs cluster radius

Figure 7 compares two different clustering approaches for bike GPS destination data:

- the method "single" specifies that the distance between two clusters is the minimum distance between a point in the first cluster and a point in the second cluster.
- the method "complete" uses the largest distance between a point in one cluster and a point in the other cluster.

It can be seen that the method "single" is more robust to cluster radius, forming a consistent number of clusters for a range of radii from 80 to 190 m, whereas the "complete" method only stabilises and agrees with this value between 150 and 190 m. We thus have selected the "single" clustering method with a radius of 120m to perform clustering.

Examples of both single and complete clusters for 50, 120 and 200m distances are shown in Figure 8 to Figure 10. It can be seen that the *single* approach always finds one cluster for the home and work destinations, ignoring the single point away from the home location. The *complete* approach finds 4 clusters for work and 3 for home at 50m, 2 for work and 2 for home at 120m, and one for home and work at 200m

This is a relatively simplistic approach – a more complete analysis would use a similar analysis based on accuracies obtained from each individual measurement device, as errors will vary depending on GPS receiver hardware and characteristics of the urban environment.

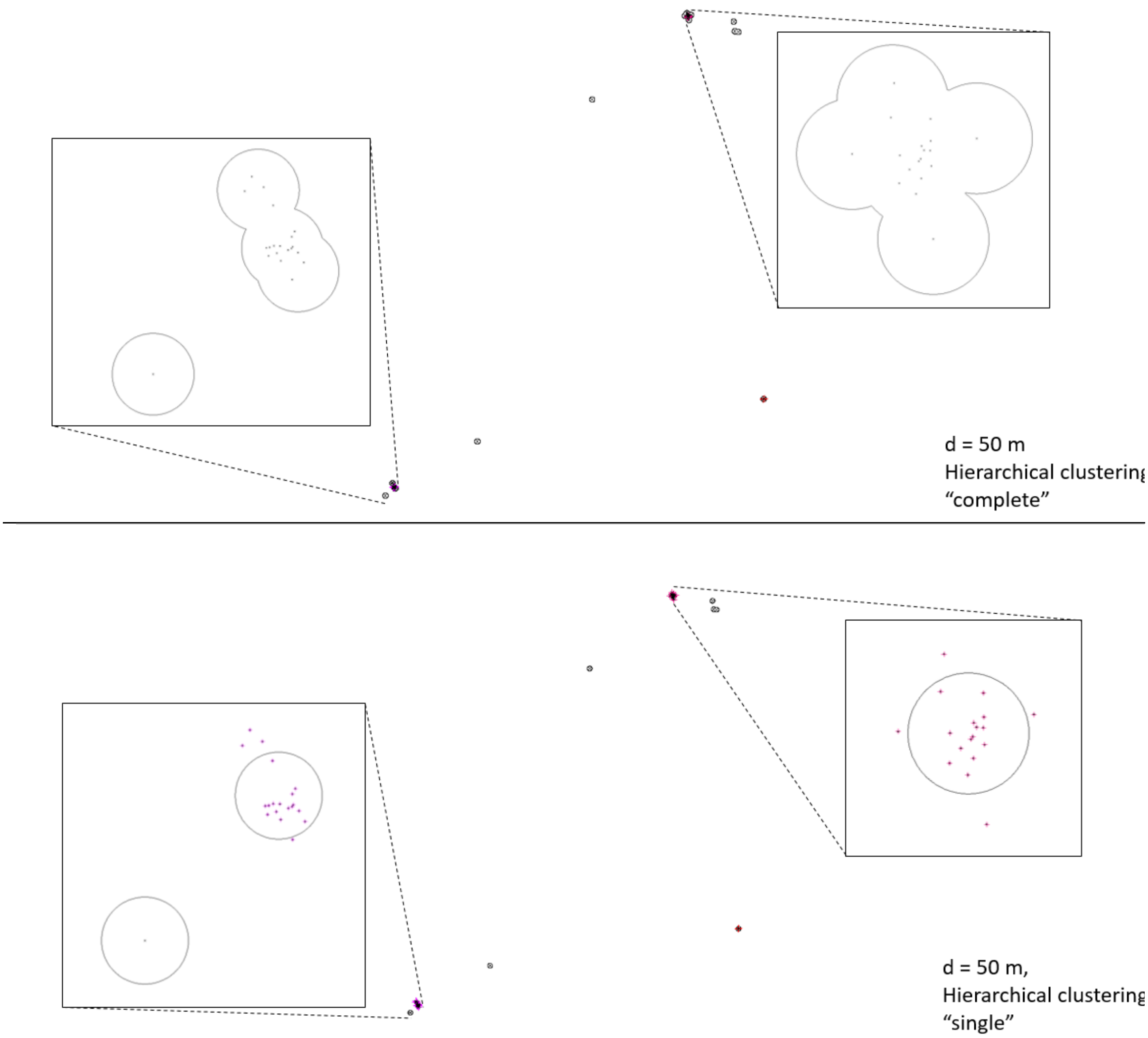


Figure 8: clusters of 50m, *complete* top, *single* bottom. Home (bottom left) and work (top right) locations shown enlarged.

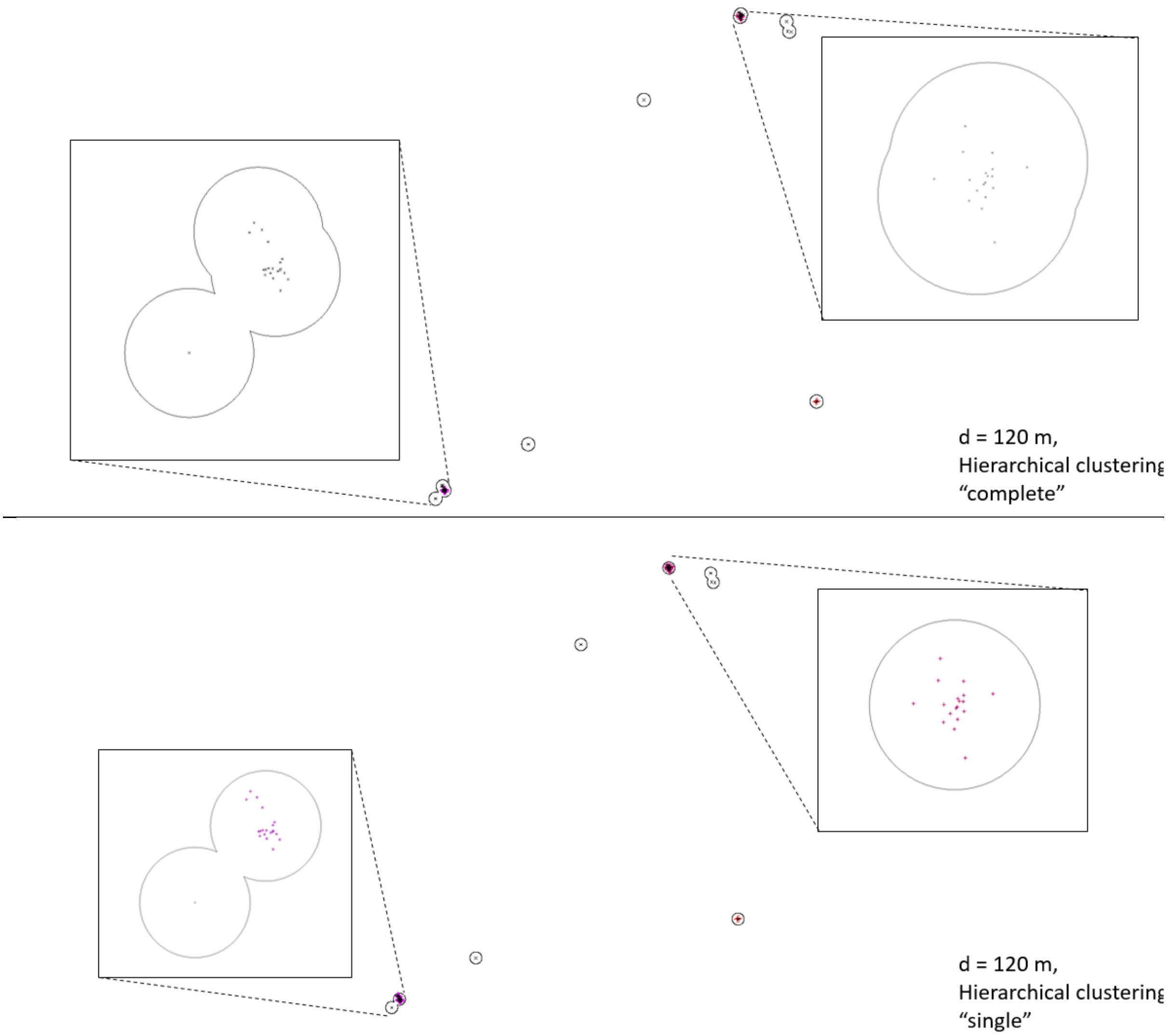


Figure 9: clusters of 120m, *complete* top, *single* bottom. Home (bottom left) and work (top right) locations shown enlarged.

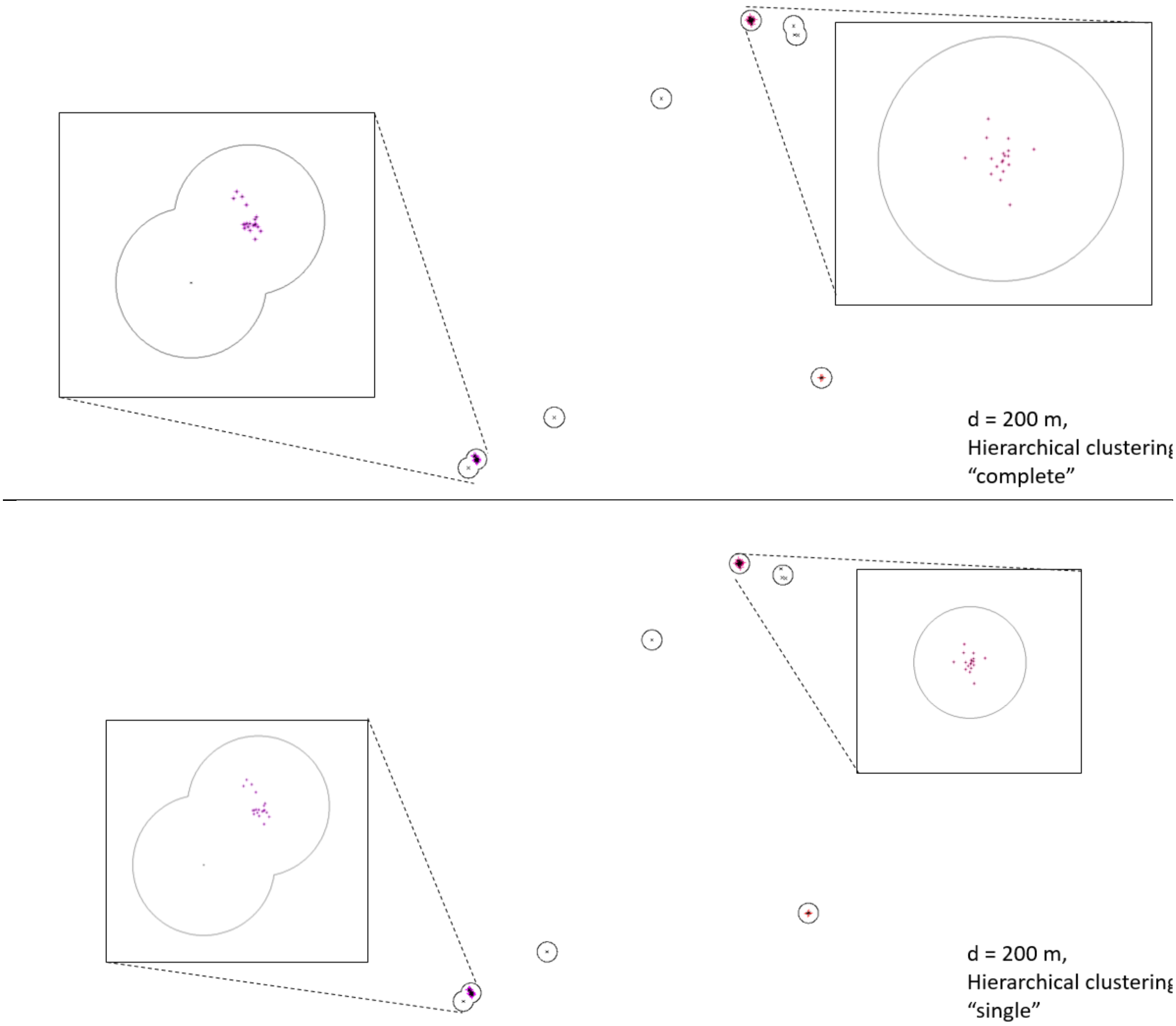


Figure 10: clusters of 200m, *complete* top, *single* bottom. Home (bottom left) and work (top right) locations shown enlarged.

7.4. Activity filtering

Whilst phones provide some estimate of activity type, the reliability of this varies between phone models, and between Android and iOS, with iOS often not recognising cycling.

A journey classification algorithm is currently being developed, using journey speed characteristics as well as phone activity estimation. Some exemplar data is shown below:

Figure 11 shows walking data, it can be seen that both the start and a midpoint of the journey are affected by GPS artefact, which leads to an inflation of both distance walked and instantaneous speed. Artefact rejection, especially of outlying speed data is important prior to classification.

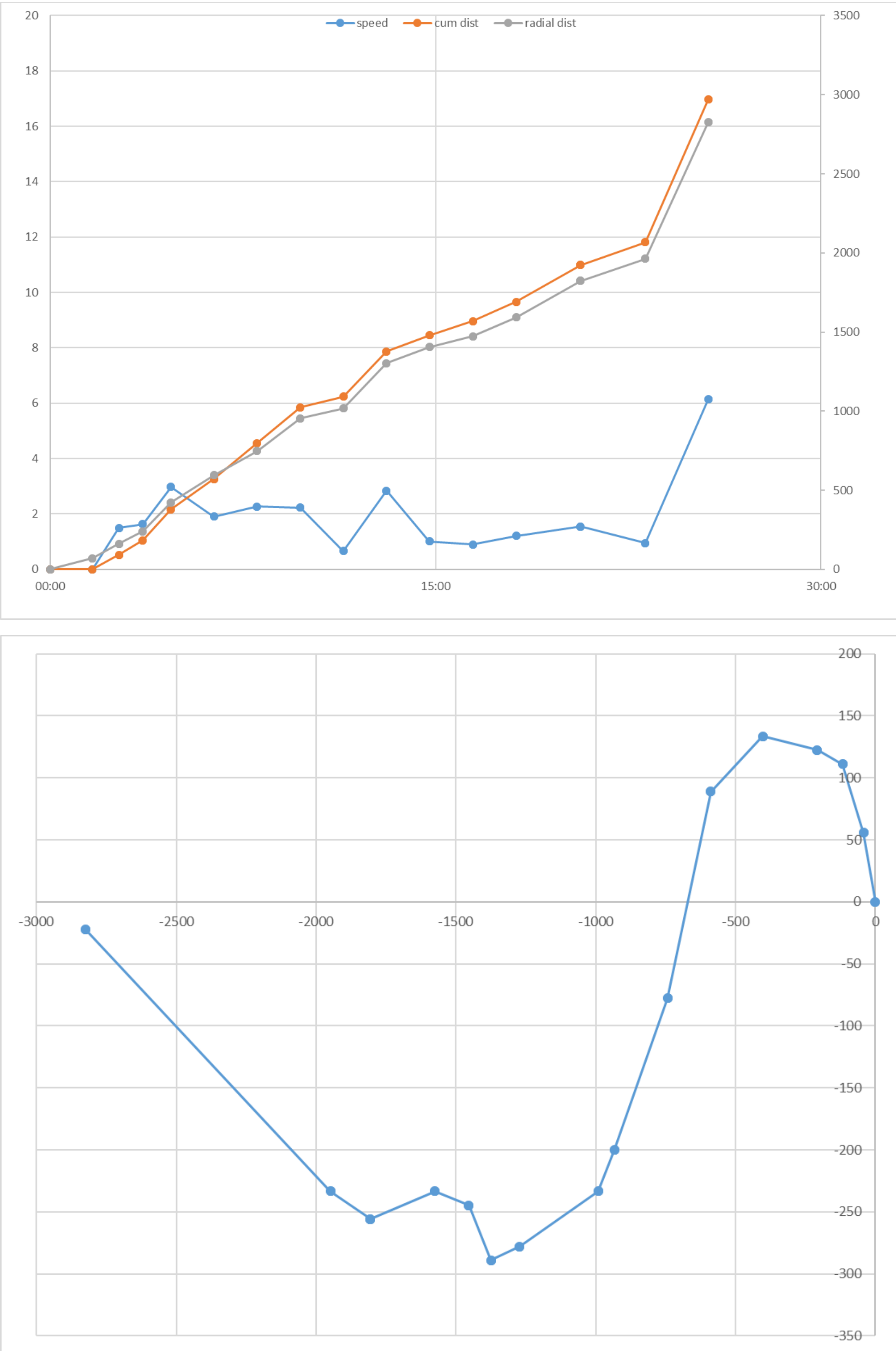


Figure 11: walking data – top: speeds (m/s) and absolute distances (m). bottom: relative position (m)

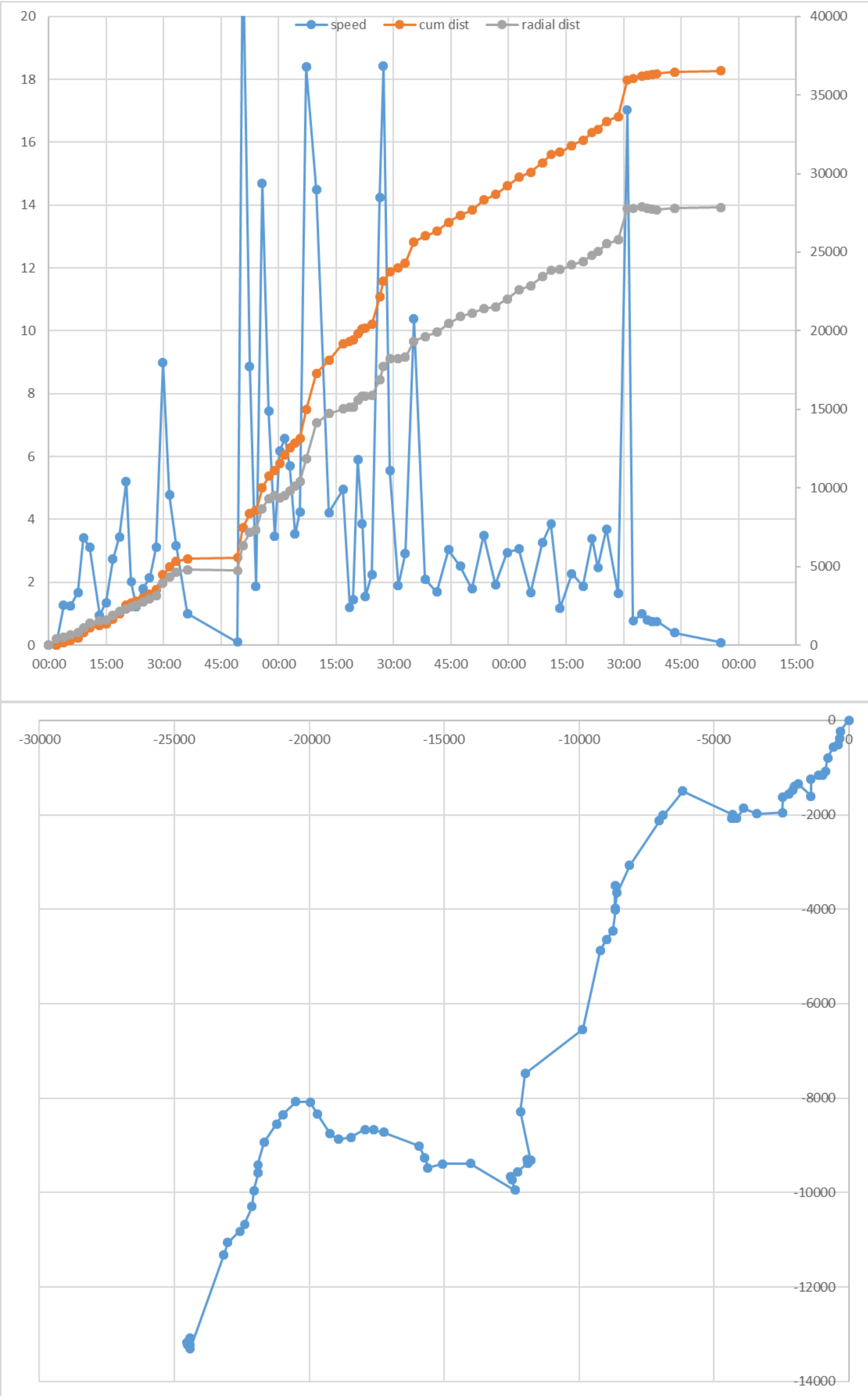


Figure 12: : cycling data - top: speeds (m/s) and absolute distances (m). bottom: relative position (m)

Figure 12 shows cycling data from a long journey with up and down hills and including both rural and urban cycling. The mean speed of 3.4 m/s indicates not walking, and the high peak values indicate cycling rather than running.

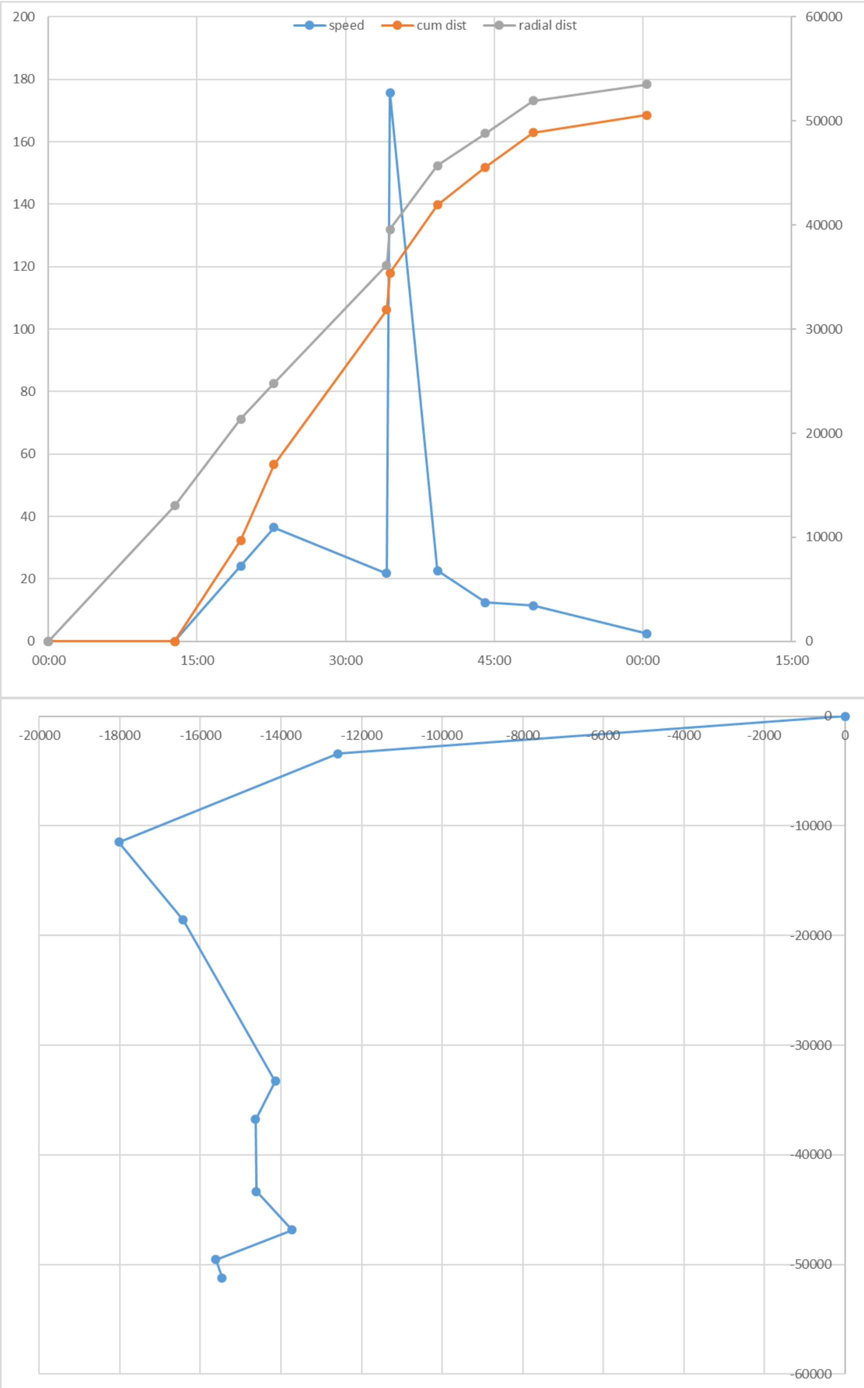
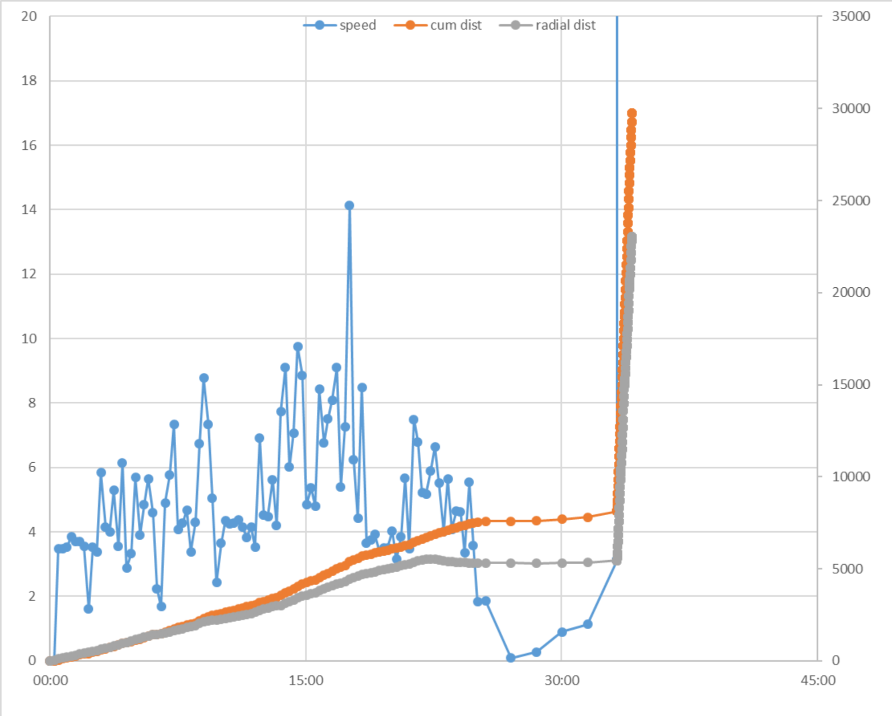


Figure 13: driving data - top: speeds (m/s) and absolute distances (m). bottom: relative position (m)

Figure 13 shows driving data. Again, an artefact is apparent with a speed of over 100 m/s. The remaining speed points show speeds characteristic of extra-urban and urban driving.



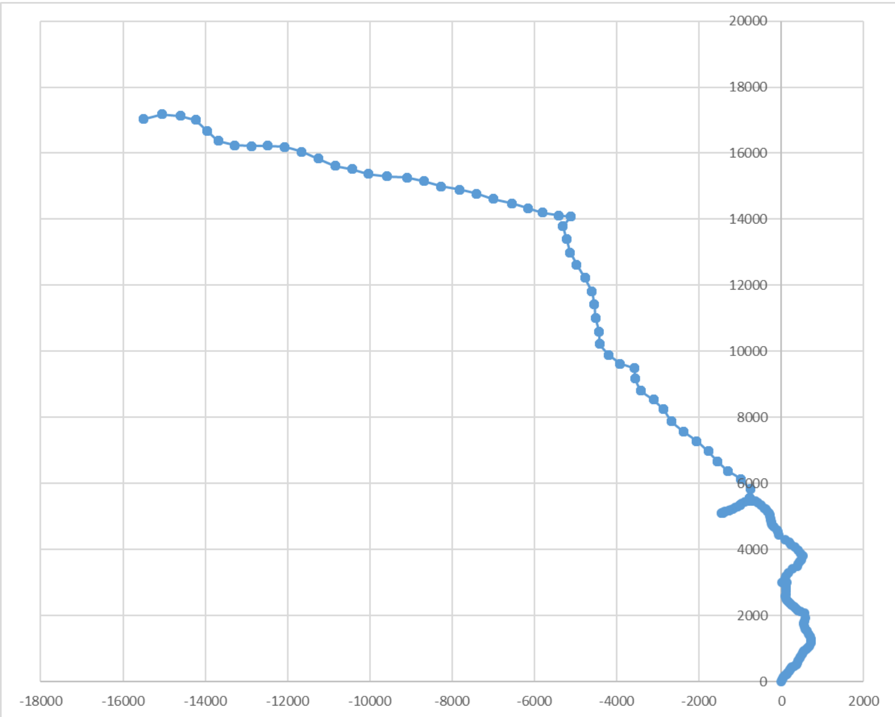


Figure 14: cycling data with artefact - top: speeds (m/s) and absolute distances (m). bottom: relative position (m)

Figure 14 shows cycling data. An artefact is apparent in the later part of the trace (after a break) that leads to return of correct positions, but with mislabelled timestamps, hence the instantaneous speeds are estimated to be far too high. Some other traces have similar artefacts, and work is currently ongoing to detect and correct artefacts such as these.

7.5. Behaviour clustering

Behavioural clustering approaches were investigated on the parkrun UK database of runs, with a view to determining clusters in patterns of attendance. There were 13 million individual runs in the database from 1.2 million runners at 449 events across the UK.

A self-organising map (SOM) was created to predict the likelihood of running in the following year, based on the number of runs in the current year. Self-organising maps are appropriate, as they are a form of unsupervised learning: that is, classes are determined from characteristics inherent in the data, rather than using pre-existing class labels. Self-organising maps consist of a projection from the multi-dimensional input space onto a two-dimensional map consisting of neurons aligned on a hexagonal grid. The mapping is adjusted in a way that preserves topographical features, keeping samples that are close in N dimensional space also nearby in the 2 dimensional space.

The 6 SOM dimensions used were runs in years 1, 2 and 3, and time between runs in years 1, 2 and 3. There were 2000 kernel points (neurons) in the self-organising map. Kernels with low hits were pruned, and then k-means was used to identify clusters. The U-matrix shows distances between adjacent kernels and is used to identify clusters in the data.

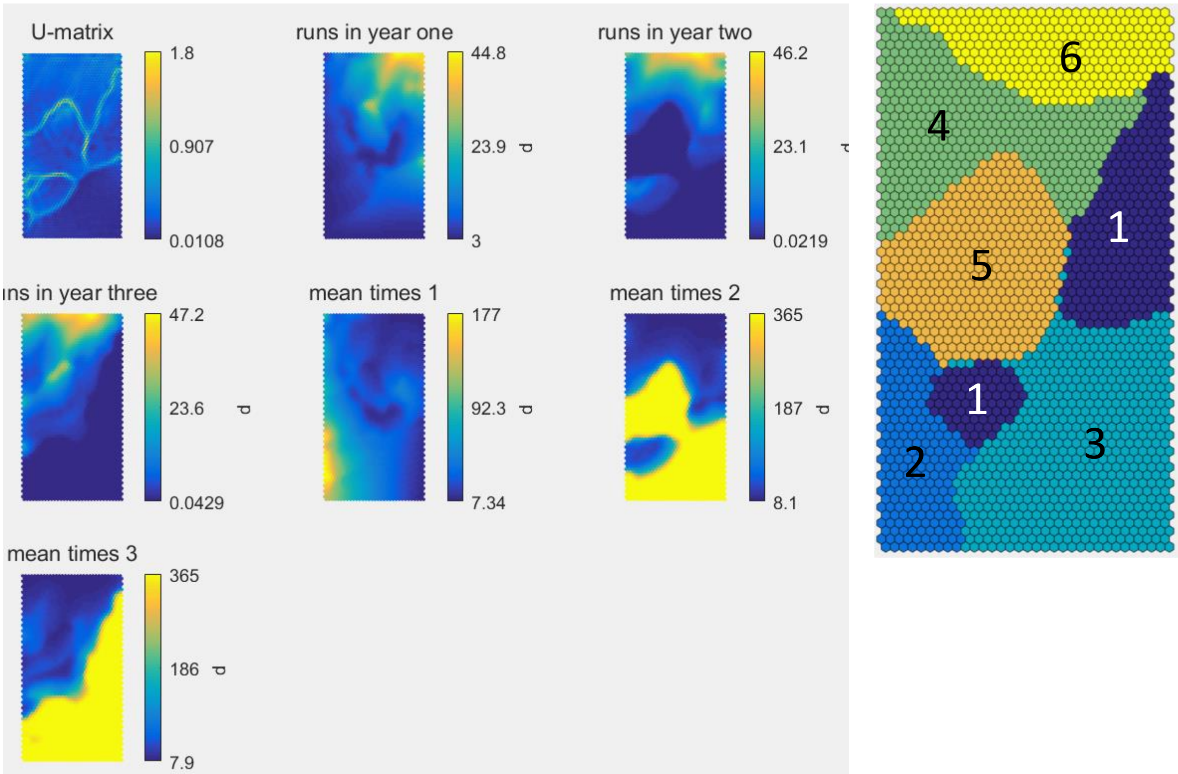


Figure 15 clustering for parkrun attendance data. The U-matrix shows a 2-dimensional view of the distance (in 6 dimensions) between neighbouring neurons in the map. Larger values show greater distances between neurons, and are indicative of clusters in the underlying data. Six detected clusters are shown on the right. Attendance behaviour across years 1-3 can be seen to vary across different areas of the map.

The characteristics of each of the identified clusters with respect to the number of runs completed per year is shown in Figure 16.

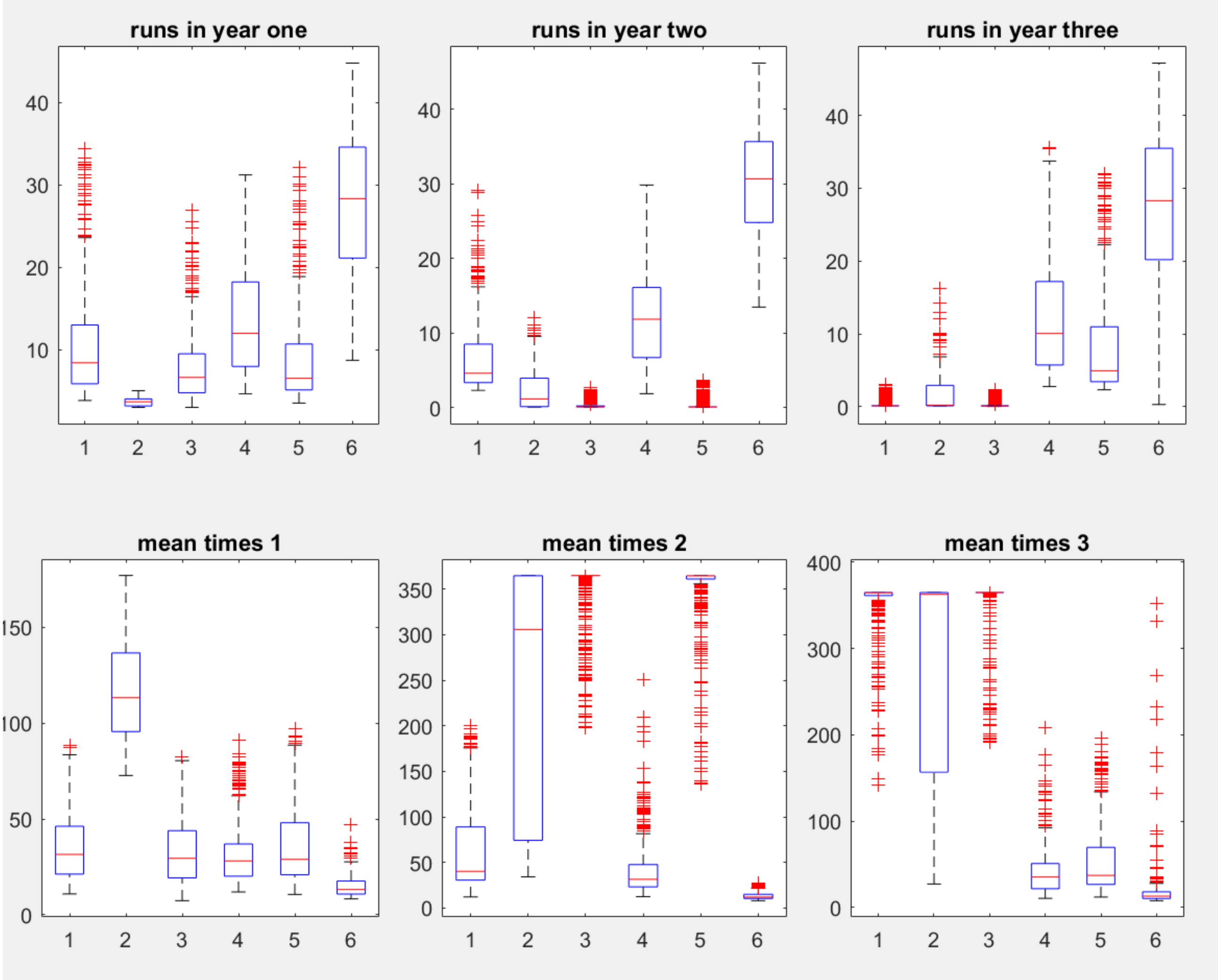


Figure 16: characteristics for the clusters shown previously.

Although these data are low dimensional and not directly analogous to the SETA movement data, the approach shows that behavioural clusters can be identified in data of similar magnitude (temporal and numerical) to the data sets expected within SETA. This clustering approach will be directly usable when the main SETA data are available.

8. Conclusions and further work

The rationale for modelling non-motorised transport has been explored, and many of the relevant determinants have been identified. A framework for modelling has been identified and work is continuing on filtering of activity type and journey type with preliminary data. Behavioural clusters will be used as proxies for individual characteristics.

Some issues remain in reducing artefacts within the journeys databases to allow accurate characterisation and mapping of non-motorised travel.

Work will continue on preparing the tools and auxiliary datasources, prior to availability of the full SETA dataset.

Once SETA data are available, the full SETA non-motorised transport model will be developed and reported in deliverable 4.8 Final modelling of non-vehicular transport, due in project month 33.

9. References

Andersen LB. (2017) Active commuting is beneficial for health. *BMJ* ;357:j1740

Bandura A. Social foundations of thought and action: a social cognitive. Englewood Cliffs, NJ: Prentice Hall; 1986. p. 45.

Bauman, A., Rissel, C., Garrard, J., Ker, I., Speidel, R., & Fishman, E. (2008). Cycling: Getting Australia moving. Melbourne: Cycling Promotion Fund for the Department of Health and Ageing.

BBC (2015). Free Birmingham bike project under way. http://www.bbc.co.uk/news/uk-england-birmingham-33154673

Ball, K (2006). People, places . . . and other people? Integrating understanding of intrapersonal, social and environmental determinants of physical activity. *Journal of Science and Medicine in Sport* 9, 367-70

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46(10), 1730-1740.

Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., & Sallis, J. F. (2009). Measuring the built environment for physical activity: state of the science. *American Journal of Preventive Medicine*, 36(4), S99-S123.

Carter, P., Bodicoat, D.H., Jones, A., Khunti, K., Davies, M.J., Edwardson, C.L., Henson, J., Yates, T. and Coombes, E., (2017). The impact of neighbourhood walkability on the effectiveness of a structured education programme to increase objectively measured walking. *Journal of Public Health* 1-8.

Celis-Morales, C.A., Lyall, D.M., Welsh, P., Anderson, J., Steell, L., Guo, Y., Maldonado, R., Mackay, D.F., Pell, J.P., Sattar, N. and Gill, J.M., (2017). Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study. *bmj*, 357, p.j1456.Clark AF,

Scott DM (2013) Does the social environment influence active travel? An investigation of walking in Hamilton, Canada. *Journal of Transport Geography* 31 278–85

Cooper, C., & Chiaradia, A. J. (2015). sDNA: how and why we reinvented Spatial Network Analysis for health, economics and active modes of transport (research report).

Cooper, C. H. (2017). Using spatial network analysis to model pedal cycle flows, risk and mode choice. *Journal of Transport Geography*, 58, 157-165.

Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S., Ogilvie, D., Petticrew, M., Reeves, B., Sutton, M. and Thompson, S., (2012). Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *Journal of Epidemiology and Community Health*, pp.jech-2011.

Department of Health, 2011 *Start active, stay active: A report on physical activity for health from the four home countries Chief Medical Officers:*

EC (2009) Action Plan on Urban Mobility (http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52009DC0490&from=EN)

Elvik, R., & Bjørnskau, T. (2017). Safety-in-numbers: a systematic review and meta-analysis of evidence. *Safety Science*, 92, 274-282.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American planning association*, 76(3), 265-294.

Frank, L.D., Schmid, T.L., Sallis, J.F., Chapman, J., Saelens, B.E., (2005). Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. *American Journal of Preventive Medicine* 28, 117–125.

Goodman, A., Sahlqvist, S., & Ogilvie, D. (2014). New walking and cycling routes and increased physical activity: one-and 2-year findings from the UK iConnect Study. *American Journal of Public Health*, 104(9), e38-e46.

King, D. M., & Jacobson, S. H. (2017). What is driving obesity? A review on the connections between obesity and motorized transportation. *Current obesity reports*, 6(1), 3-9.

Lovelace, R., 2015. Crowd sourced vs centralised data for transport planning: a case study of bicycle path data in the UK. *GIS Research UK* (GISRUK). Leeds http://leeds.gisruk.org/abstracts/GISRUK2015_submission_71.pdf.

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*, 6(1), 42.

Murtagh, F. (1985). Multidimensional Clustering Algorithms, in *COMPSTAT Lectures 4*. Wuerzburg: Physica-Verlag

Oliver, M., Badland, H., Mavoa, S., Duncan, M. J., & Duncan, S. (2010). Combining GPS, GIS, and accelerometry: methodological issues in the assessment of location and intensity of travel behaviors. *Journal of Physical Activity and Health*, 7(1), 102-108.

Rivis, A., & Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. *Current Psychology*, 22(3), 218-233

Rojas-Rueda, D., de Nazelle, A., Tainio, M., & Nieuwenhuijsen, M. J. (2011). The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *Bmj*, 343, d4521.

Rowe, K., Shilbury, D., Ferkins, L., & Hinckson, E. (2013). Sport development and physical activity promotion: An integrated model to enhance collaboration and understanding. *Sport Management Review*, 16(3), 364-377.

Sallis, J.F., Spoon, C., Cavill, N., Engelberg, J.K., Gebel, K., Parker, M., Thornton, C.M., Lou, D., Wilson, A.L., Cutter, C.L. and Ding, D. (2015). Co-benefits of designing communities for active living: an exploration of literature. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 30.

Saunders, L. E., Green, J. M., Petticrew, M. P., Steinbach, R., & Roberts, H. (2013). What are the health benefits of active travel? A systematic review of trials and cohort studies. *PLoS One*, 8(8), e69912.

SETA (July 2016) *Deliverable 4.1- Exploring prediction perspectives*

Smith J & Blewitt, R (eds) (2010) Traffic Modelling Guidelines TfL Traffic Manager and Network Performance best practice v3.0. http://content.tfl.gov.uk/traffic-modelling-guidelines.pdf

Trost SG, Owen N, Bauman A, Sallis J, Brown W. (2002) Correlates of adults' participation in physical activity: review and update. *Med Sci Sport Exerc*;34(12):1996—2001.

Van Cauwenberg, J., De Bourdeaudhuij, I., De Meester, F., Van Dyck, D., Salmon, J., Clarys, P., & Deforche, B. (2011). Relationship between the physical environment and physical activity in older adults: a systematic review. *Health & Place*, 17(2), 458-469.

Workout-UK, (2016) - Major activity programme launches in Sheffield (http://www.workout-uk.co.uk/news/807/Major-activity-programme-launches-in-Sheffield)