Aalborg Universitet

Distinguishing movement from stays during continual GPS tracking

Danish working title: Raffinering af data fra GPS-baserede undersøgelser

Jensen, Anders Sorgenfri; Bro, Peter; Harder, Henrik; Hansen, Jakob Hjorth; Tradisauskas, Nerius

Distinguishing movement from stays during continual GPS tracking

(Danish working title: Raffinering af data fra GPS-baserede undersøgelser)

Anders Sorgenfri
Peter Bro
Henrik Harder
Jakob Hjorth Hansen
Nerius Tradisauskas

Department of Architecture and Design
Aalborg University, Denmark

During the past couple of years, the research group "Diverse Urban Spaces" has performed a thorough survey of how the Danish city of Aalborg is being used by a certain population group. This group consists of young people at age 17-23 years whose main occupation is high school or equivalent level of education.

The core of the survey revolved around each participant (henceforth referred to as respondent) carrying a pocket-sized GPS receiver for a period of seven days. The GPS receiver would record the position of the respondent approximately every 8th second. With a sample of 169 hand-picked respondents selected from a total group of 212 respondents, Diverse Urban Spaces gathered a vast amount of geodata which could be used to analyse popular city spaces, plazas and buildings as well as pointing out which roads and streets the group would use most frequently. [Harder et. al page 5]

However, the nature of everyday life involves a relatively high amount of immobility compared to movement. Attempting to analyse how this setup of respondents use the city becomes problematic when the gathered geodata consists of both the 1-2 hours each respondent spend on travelling through the city, and the remaining 22 hours of the day which the respondent would spend stationary at certain locations important to the respondent such as school, home address, work address etc. It would be much more convenient if the geodata could be divided into two datasets consisting of movement and stays respectively, so stay-related analyses such as pinpointing popular city squares wouldn't be affected by noise from movement data.

This paper aims to describe the technique developed by Diverse Urban Spaces to counteract the above mentioned inconvenience by splitting the dataset into movement and stays. The paper will explain how the technique was commenced, how it works and its level of quality.

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Results of the tracking

As mentioned in the abstract, the GPS receiver used in the survey records its position roughly every 8th second when turned on and transmits this information to a database. Every logged position is then stored as a record in the database. The result is a large table containing a total of 9.785.201 records.

The output of a single GPS receiver can be visualised using the georeference stored in the database. As illustrated on Figure 1, every record represents a point in space at a specific time. In the example, we're probably monitoring a respondent's trip back home from school.
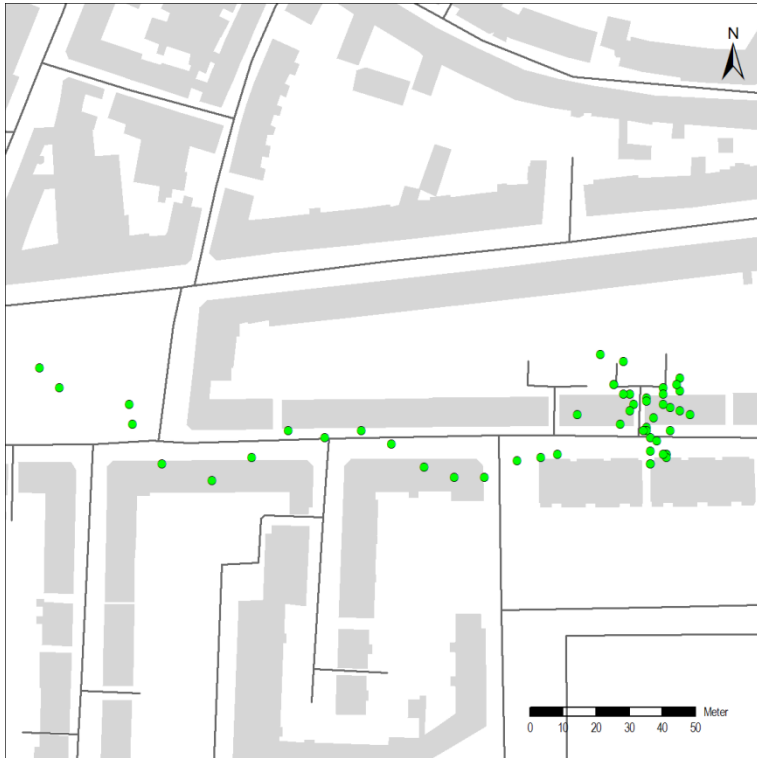
Figure 1 A visualisation of a sample from the output generated by a single GPS receiver. The sample contains 61 points which is equivalent to a time span of approximately 8 minutes and 8 seconds.

Furthermore, the visualisation illustrated on Figure 1 shows how movement follows a somewhat linear pattern along the road, whilst points recorded during the respondent's stay at his home results in inaccurate positioning which is recorded randomly in close proximity to the respondent's actual location. This scattering pattern is fundamental for the trip/stay-defining procedure, which will be explained in the following chapter.

Figure 1 shows the raw output of a single GPS receiver visualised as points for each record. This visualisation, however, is only appropriate when dealing with a single respondent and browsing a small scaled map. Figure 2 is an illustration of the entire dataset, containing several million records and Figure 3 shows the base map of central Aalborg City beneath the dataset.

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University
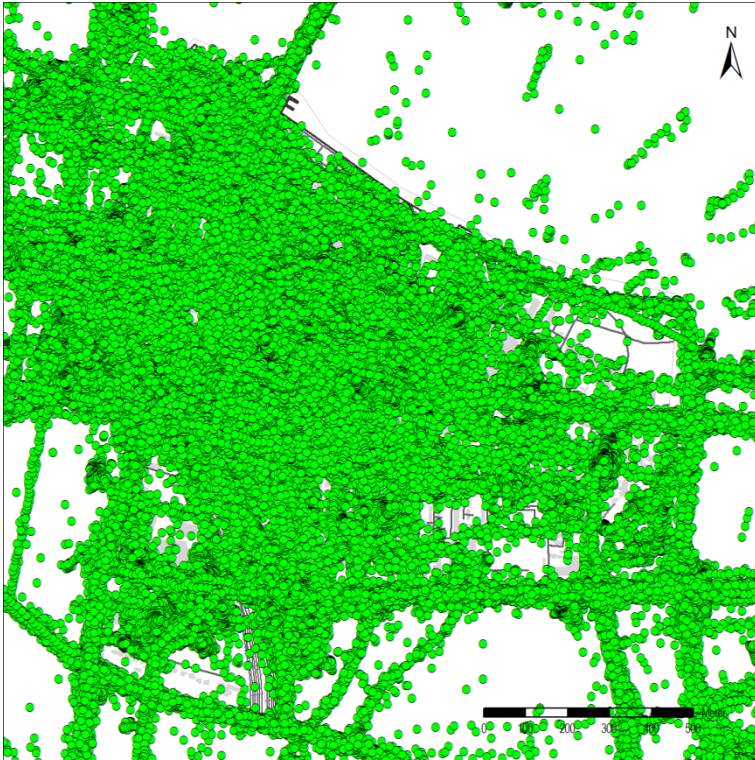
Figure 2 The entire dataset's records visualised as points

Figure 3 The base map of the central urban area of Aalborg City used in Figure 2

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

It is quite obvious, that this method of visualisation isn't well suited for portraying an overall pattern as to how the respondents use the city as a group since the entire map is completely covered by point records.

To counteract this, the visualisation was changed to an aggregation where all records are summed up upon a grid of squares with dimensions of 5 x 5 meters for map scale 1:5000, 50 x 50 for map scale 1:50000 and 250 x 250 for map scale 1:250000. The sum is then used to reflect how populated each square was during the survey. Figure 4 illustrates the result of performing the aggregation process on the sample dataset used in Figure 1.

Figure 4 The sample dataset used in Figure 1 as seen after the aggregation has been applied

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Note that all squares have the same colour. This is due to the fact, that the sample dataset is very small and thus, only a maximum of two points lie within the same square. In contrast, the aggregation performed on the entire dataset can be seen on Figure 5.

Legend:
- 1 sec - 20 sec
- 20 sec - 40 sec
- 40 sec - 60 sec
- 60 sec - 1 min 40 sec
- 1 min 40 sec - 2 min 40 sec
- 2 min 40 sec - 3 min 20 sec
- 3 min 20 sec - 7 min
- 7 min - 11 min 20 sec
- 11 min 20 sec - 18 min 20 sec
- 18 min 20 sec - 29 min 40 sec
- 29 min 40 sec - 48 min
- 48 min +

Meter
0   100   200   300   400   500
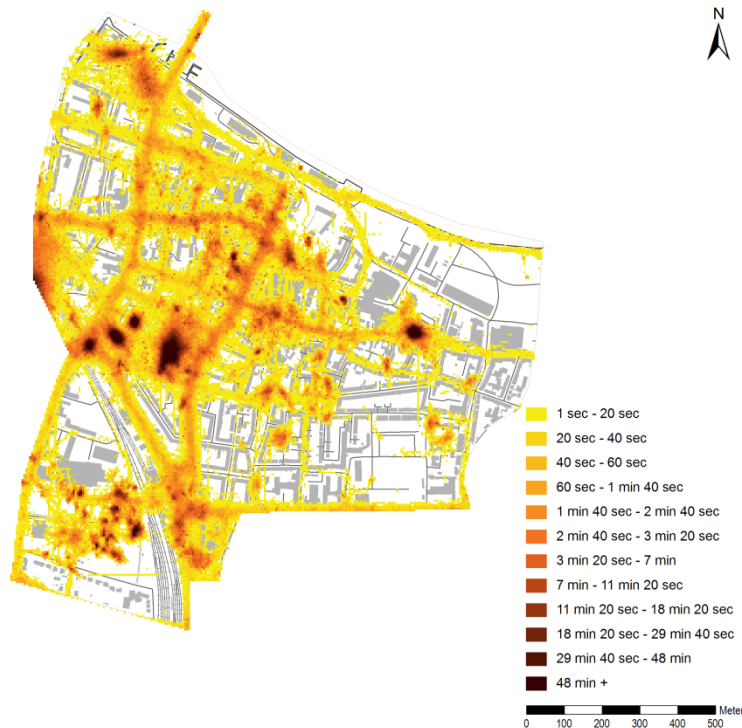
Figure 5 The entire dataset as seen after the aggregation process

Figure 5 is a decent visualisation in its own since it displays several hotspots in the city, and it is still possible to see which roads and streets are most frequently used. The visualisation has drawbacks though. It is impossible to define what that source of the hotspots is, since the map in most of these areas is covered by squares. Likewise, it isn't possible to get a nuanced interpretation of just how much more populated a certain street is compared to another since the population colour scale has to display a nuanced picture of the hotspots, making the colour gradually darker the closer the squares get to the core of the hotspot.

The above mentioned problems with the final visualisation of how the respondent use the city as a group is the motivation that led to the creation of the trip/stay-recognition procedure. The following chapter will explain how the initial version of the procedure came to be, and how it worked.

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Page 5

The initial version of the procedure

As mentioned in the previous chapter, the observation which started the development of the procedure, was the fact that the GPS receiver would have trouble recording its actual position as soon as it would be carried inside a building. Since the result of recording positions inside a building was a scattering pattern of points surrounding the actual position, this sparked the idea, that there was a connection between this pattern and the respondent remaining stationary at his home/school/workplace. The idea also involved an anticipation of different attribute values of points recorded along roads, paths etc compared to points recorded during the scattering phase.

These attributes would be 2 values, which the GPS receiver records alongside the georeference – speed and cardinal direction. An example of how these two attributes can be visualised is shown on Figure 6.
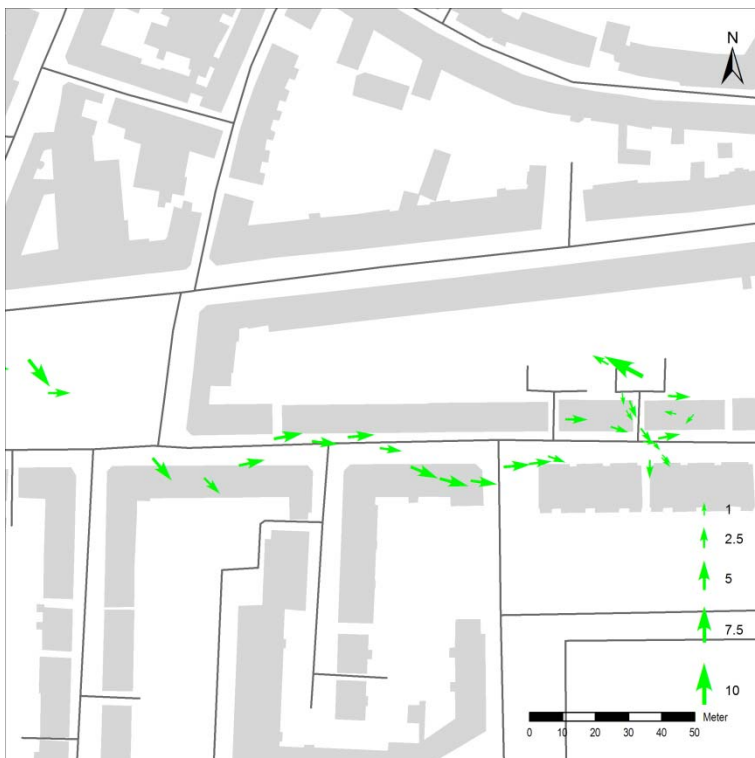
Figure 6 An illustration as to how the attributes speed and cardinal direction can be visualised. The cardinal direction dictates the direction of each point, and the speed value determines the size. The points used in the figure are the same points used in Figure 1.

As Figure 6 exemplifies, every record can be shaped as an arrow pointing in the cardinal direction as recorded by the GPS receiver, and the size of the arrow is proportional to the speed logged by the receiver.

Note that there is little change in direction from point to point when the respondent is on the move and the speed recorded rarely goes below 5 km/h. On the other hand, the scattering points recorded when the respondent is located inside a building points in random directions and there's quite a big change in direction between subsequent points. Furthermore, a lot of the points recorded during a stay have a very low speed value – usually in the range 0 – 2 km/h.

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Page 6

This observation consisting of a continuous, dramatically change of direction and a similar continuous low speed value over a series of points recorded during the respondent's stays as opposed to a minor change in direction and a higher speed value over a series of points recorded during the respondent's trips is the backbone of the first procedure which would define a point as being either a trip point or a stay point.

Mechanisms of the initial version of the procedure

The procedure is basically an algorithm which scans the entire dataset and evaluates every record. Based on the attributes of each point and the attributes of adjacent points, the procedure will return a certain value. This value can later be used to determine a point as either a trip or a stay point depending on what would be most likely according to the performed calculations.

The procedure works like an expanded database cursor which operates with two parameters: The focal point and the window border value. Table 1 and Figure 7 shows both parameters. The focal point P in this case is the point with the ID number 15 highlighted with red. The window border value is 5 in this case, which means that the 5 points back and forth in the record list will be used for the calculations. These points are highlighted with blue.

Point ID	Speed	Direction	Focal point relation
7	4	139	-8
8	3	144	-7
9	4	103	-6
10	7	142	-5
11	3	90	-4
12	0	177	-3
13	0	157	-2
14	5	140	-1
15	3	135	P
16	4	78	1
17	5	79	2
18	4	97	3
19	4	85	4
20	3	98	5
21	5	113	6
22	5	105	7
23	4	98	8
24	4	85	9
25	3	84	10
26	2	110	11

Table 1 Database table view of the focal point, the points residing within the window border and points which won't be used in the calculations for the focal point P

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Page 7
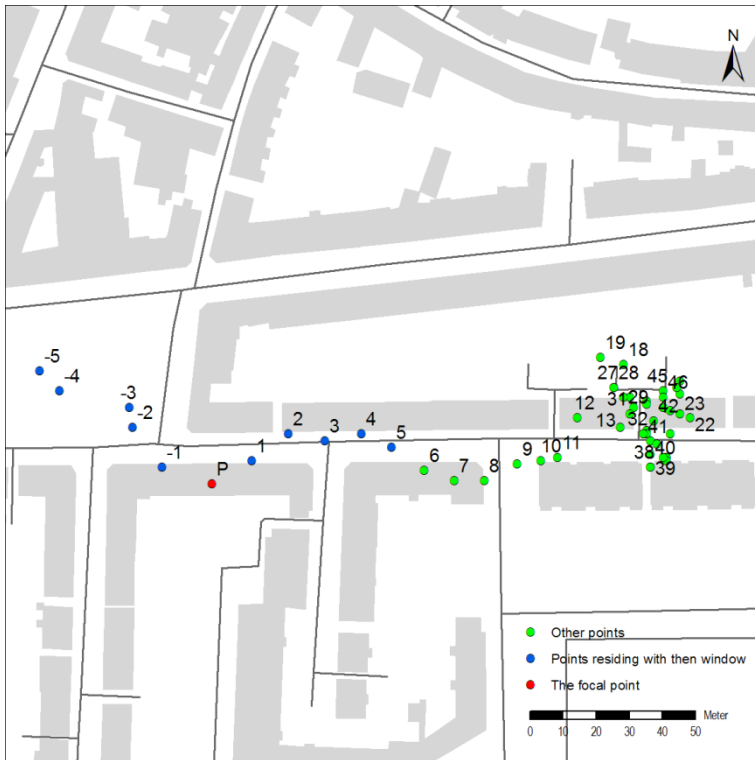
Figure 7 Visualisation of the database table containing the focal point, points residing within the window order and points which won't be used in the calculations for the focal point P

Now, a series of intermediate sub calculations are performed upon each of the attributes speed and direction. These values aren't stored anywhere, but Table 2 illustrates how they are derived – in the case, the attribute is direction.

Point ID	Speed	Direction	Focal point relation	Direction difference
7	4	139	-8	
8	3	144	-7	
9	4	103	-6	
10	7	142	-5	
11	3	90	-4	52
12	0	177	-3	87
13	0	157	-2	20
14	5	140	-1	17
15	3	135	P	62
16	4	78	1	1
17	5	79	2	18
18	4	97	3	12
19	4	85	4	13
20	3	98	5	
21	5	113	6	

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

22	5	105	7	
23	4	98	8	
24	4	85	9	
25	3	84	10	
26	2	110	11	

Table 2 The same database table as in Table 1 including the temporary field "Direction difference"

As Table 2 implies, the difference in degrees between two adjacent points' direction value is calculated for every point residing within the window border. The value is stored in the record which has the lowest distance from the focal point in terms of row position. The value stored in the same record as the focal point is special since it contains a sum of the difference between the direction of the focal point and the -1 point and the difference between the direction of the focal point and the +1 point.

After these temporary values have been calculated, the sum of all these differences is calculated and stored in the new, permanent attribute field called "Direction change sum", which is illustrated in Table 3.

Point ID	Speed	Direction	Focal point relation	Direction change sum
7	4	139	-8	
8	3	144	-7	
9	4	103	-6	
10	7	142	-5	
11	3	90	-4	
12	0	177	-3	
13	0	157	-2	
14	5	140	-1	
15	3	135	P	282
16	4	78	1	
17	5	79	2	
18	4	97	3	
19	4	85	4	
20	3	98	5	
21	5	113	6	
22	5	105	7	
23	4	98	8	
24	4	85	9	
25	3	84	10	
26	2	110	11	

Table 3 The result of the calculations for the focal point P – the Direction change sum value, which is equal to the sum of all the intermediate direction differences shown in Table 2.

After this value has been calculated, the cursor consisting of the focal and the window border will jump one record downwards and the whole process is repeated for this new focal point. The process is iterated for every record in the entire dataset. Besides the attribute containing the sum of the direction change, the

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Page 9

calculations are performed in a similar manner for the speed attribute and the distance between two adjacent points, which can be derived through applying the Pythagorean theorem on the points' georeferences.

Thus, the output of the procedure is 3 attributes: DirsumX, SpeedsumX and DistsumX where X is the size of the border window – in the example illustrated above, X would be 5.

Applying the initial version of the procedure

Since the calculations described in the previous chapter returns values in a certain range for trip points, and values in another range for stay points, the next step was to apply the procedure on a dataset where every point was predefined as belonging to either a trip or a stay.

The procedure was applied to a dataset consisting of 10.000 points of each type. This dataset was created by visualising the whole dataset according to its georeferences, and judging by the points' location in space compared to each other, and compared to a base map consisting of roads and buildings, an employee would manually mark points as either a trip or a stay point until a total of 10.000 points of each type were marked.

The result of applying the procedure on the predefined dataset was a series of graphs, containing a plot of trip points and stay points for each survey. An example of these graphs is shown on Figure 8.
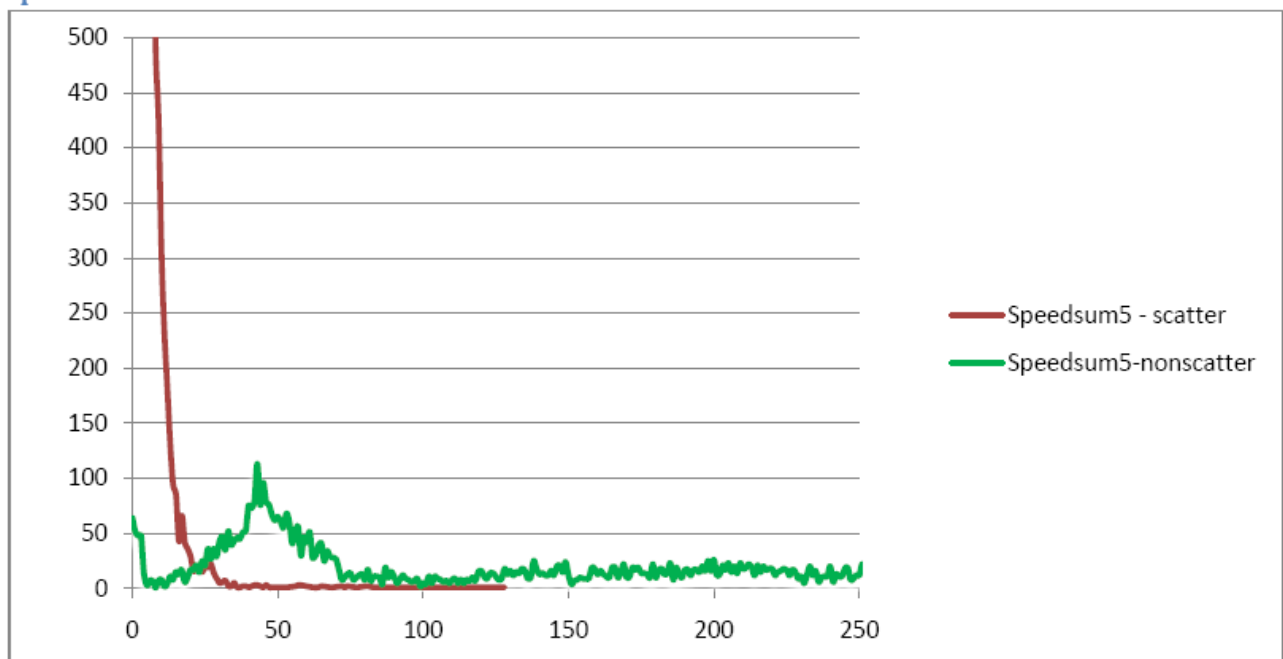

Figure 8 A graph which illustrates calculated Speedsum5 values and number of occurrences of these values for trip points and stay points respectively

The x-axis would contain the attribute values (in the case of Figure 8, the Speedsum5 value), and the y-axis contains the number of occurrences for these values for trip points (red) and stay points (green).

A complete list of all these graphs can be found in [Jensen et. al.].

As Figure 8 implies, there's an intersection between the two plots at approximately around the Speedsum5 value 25. For records with a Speedsum5 value lower than 25, there's a much greater likelihood that the point is a stay point. Likewise, there's a greater likelihood that records with a Speedsum5 value greater than 25 is a trip point.

This information in conjunction with specific Dirsum and Distsum values led to a query, which would mark approximately 70 % of the points correctly.

Inconveniences of the initial version of the procedure

The first version of the procedure, as described by the previous chapters, was a milestone since it managed to automatically divide a good portion of a dataset into trip points and stay points respectively. Approximately 30 % of the points were erroneously marked, however, which gave incentives to review the procedure for weaknesses, which could be the source for the wrongly marked points.

The main weakness turned out to be the static nature of the window border. Since the procedure will always include the specified amount of points adjacent to the focal point, certain factors aren't taken into account. One of these would be the difference in time between two subsequent points. If the respondent turned the GPS receiver off or it ran out of batteries and later turned it back on, the procedure wouldn't notice.

Point ID	Timestamp	Focal point relation
212	03-11-2008 16:53:21	-5
213	03-11-2008 16:53:29	-4
214	03-11-2008 16:53:37	-3
215	03-11-2008 16:53:45	-2
216	03-11-2008 16:53:53	-1
217	03-11-2008 16:54:01	P
218	03-11-2008 16:54:09	1
219	03-11-2008 16:54:17	2
220	03-11-2008 16:54:25	3
221	03-11-2008 18:15:15	4
222	03-11-2008 18:15:23	5

Table 4 The procedure doesn't take great leaps in time between two records into account

This could potentially mess up the results of the calculations, since the GPS receiver could have gone off during a trip in town and then getting turned on as the respondent recharged it hours later when he's home again, resulting in a great leap in time, speed, direction and most notably distance.

Similarly, the static number of points adjacent to the focal points doesn't sense if two subsequent points were recorded by two different GPS receivers.

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Page 11

Point ID	GPS ID	Focal point relation
212	245112003	-5
213	245112003	-4
214	759112003	-3
215	759112003	-2
216	759112003	-1
217	759112003	P
218	759112003	1
219	759112003	2
220	759112003	3
221	759112003	4
222	759112003	5

Table 5 The procedure doesn't detect if there's a change in GPS receiver

This lack of source detection cause the same noise in the calculations as the lack of taking leaps in time into account.

The uncovering of these two flaws – the lack of checks for huge jumps in time and the lack of checks for two adjacent points being recorded by the same GPS receiver unit – which originated from the static window border led to the creation of the current version of the procedure.

The current version of the procedure

The current version resembles the first version in a sense, that the focal point and the border window still exists, and the output is still a calculated value based on the relationship between the points within the window.

What differentiates the current version from the older one is that instead of using a static number of adjacent points the border window is instead composed of a time variable. The procedure takes a user specified number of seconds, which determines how many points adjacent to the focal point will be used for calculation the procedures output. Table 6 illustrates how the window is defined when the procedure is run using a time variable of 35.

Point ID	Timestamp	Difference in time compared to the focal point	Point within window range
212	03-11-2008 16:53:21	-40 seconds	No
213	03-11-2008 16:53:29	-32 seconds	Yes
214	03-11-2008 16:53:37	-24 seconds	Yes
215	03-11-2008 16:53:45	-16 seconds	Yes
216	03-11-2008 16:53:53	-8 seconds	Yes
217	03-11-2008 16:54:01	P	-
218	03-11-2008 16:54:09	8 seconds	Yes

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Page 12

219	03-11-2008 16:54:17	16 seconds	Yes
220	03-11-2008 16:54:25	24 seconds	Yes
221	03-11-2008 18:15:15	1 hour 21 minutes 14 seconds	No
222	03-11-2008 18:15:23	1 hour 21 minutes 22 seconds	No

Table 6 An illustration of how the window border is defined upon the same data set as Table 4, when the current procedure is run with a user specified time variable of 35

This change in the window border definition resolves the inconvenient issue of involving points with too much difference in recorded time stamp compared to the focal point.

Technically, the check for whether a point adjacent to the focal point lies within the window border or not is evaluated through an iterative process. The procedure will first check the point closest to the focal point in terms of record location and evaluate if the difference in time between these two points is lower than the user specified time variable. If this check holds true, the point is validated and its attribute values will be used in the calculations. Afterwards, the process is repeated using the second-closest point and so on until the procedure reaches a record, where the difference in time compared to the focal point is greater than the specified time variable.

To counteract the almost identical issue with a switch in ID number of the GPS receiver unit, a similar check is applied, which evaluates whether or not a point which candidates for residing within the window border was recorded by the same GPS receiver unit as the focal point. This check is exemplified in Table 7.

Point ID	GPS ID	Point recorded by same device as focal point
212	245112003	No
213	245112003	No
214	759112003	Yes
215	759112003	Yes
216	759112003	Yes
217	759112003	-
218	759112003	Yes
219	759112003	Yes
220	759112003	Yes
221	759112003	Yes
222	759112003	Yes

Table 7 The dataset fraction of Table 5 exposed to the check for whether or not a point was recorded using the same GPS receiver as the focal point

Besides these new checks which fixes the main issue with the older procedure, a few other changes was applied.

Firstly, the output attribute values were changed from sums to averages since the window border is no longer homogeneous in size.

Secondly, the amount of attributes which are used for calculating values were expanded to also include number of satellites, horizontal delusion of precision which are all additional attributes recorded by the GPS receiver similar to speed and direction.

Lastly, two non-average calculations were added. These would be maximum distance from the focal point and number of points lying close to the focal point.
The maximum distance would be the highest measured distance between the focal point and the points residing within the window border.
The number of points lying close to the focal point would be a proximity test revolving around a user specified radius. Table 8 illustrates how this value (henceforth referred to as Distsum) is extracted when the user specified radius is 30.

Point ID	Distance from the focal point	Distsum value to be increased based on distance from the focal point	Distsum
212	82.52 m	No	
213	60.10 m	No	
214	44.32 m	No	
215	30.16 m	No	
216	14.15 m	Yes	
217	-	-	6
218	1.12 m	Yes	
219	0.98 m	Yes	
220	3.12 m	Yes	
221	2.98 m	Yes	
222	3.01 m	Yes	

Table 8 An example of how the points residing within the window border are used to calculate the Distsum value

This concludes the explanation of the current procedure. The next and last chapter will discuss the quality of the procedure. For a more thorough explanation of the rationale of this application see [Bro, Kwan, forthcoming]

Quality of the current procedure

As explained in the previous chapters, the output of the procedure is an array of new attributes derived from the attributes originally recorded by the GPS receiver. Each of these new attributes hints if a point belongs to a trip or a stay. To find out which attributes gives the best division quality, the procedure was run with various window border values and different radius settings (for the Distsum attribute) on a larger, manually predefined dataset compared to the earlier version of the procedure.

The currently deployed algorithm which executes the division makes use of just one of these attributes – the Distsum. The result is that the algorithm defines 94.8 % of the manually marked stay points correctly and 82.3 % of the manually marked trip points correctly. [Bro, Kwan, forthcoming]

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Page 14

The effects of applying the division upon the entire dataset can be seen on Figure 9, Figure 10 and Figure 11

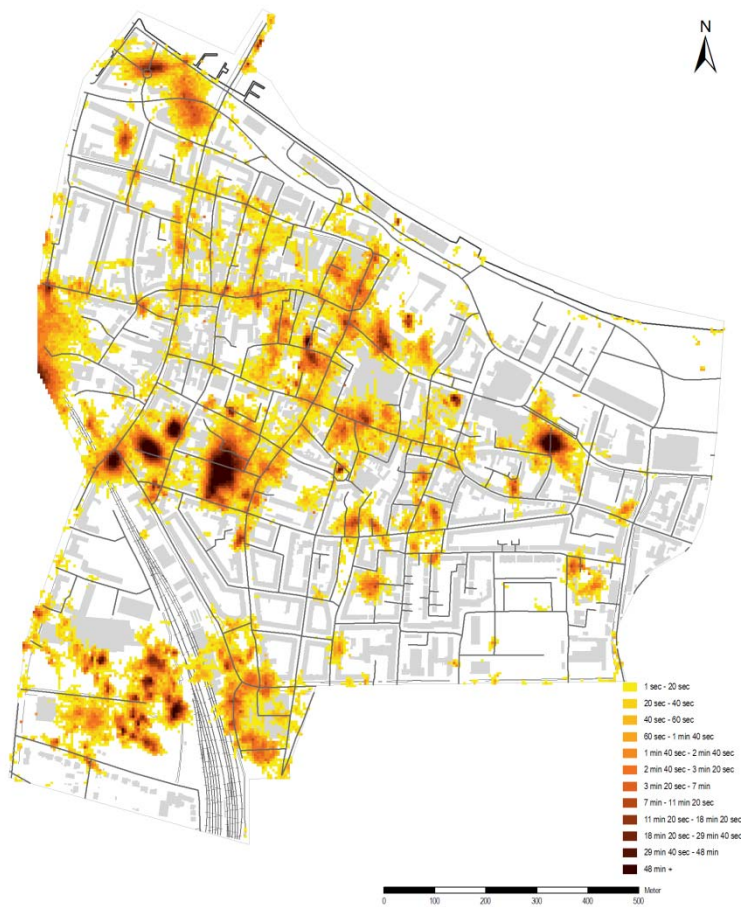**Figure 9 The entire data set before the division into trips and stays**

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Figure 10 The result of the procedure displaying only the respondents' stays

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
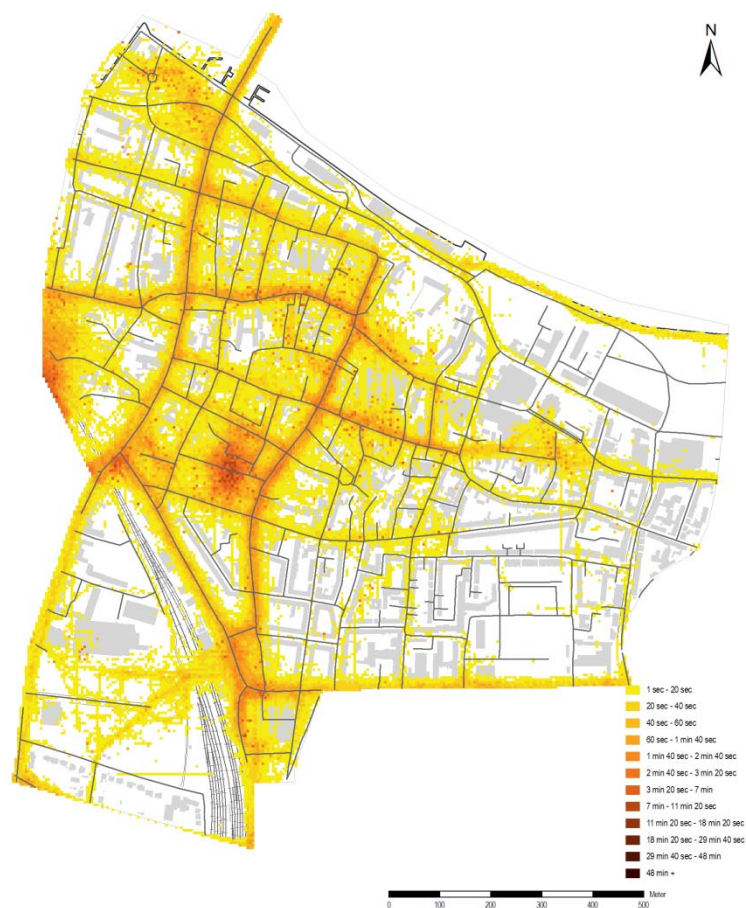Diverse Urban Spaces, Aalborg University

Figure 11 The result of the procedure displaying only the respondents' trips

The above figures show, that the procedure does quite a good job as hot spots throughout the city can be seen without the trips bluring out the base map. Likewise, a more nuanced picture of which roads and streets are used most frequently can be drawn since the extreme values stored in squares located near the respondents' schools and homes have been filtered out. A list of the amount of points used for the three visualisations can be found in Table 9.

	The entire dataset (Figure 9)	The stays dataset (Figure 10)	The trips dataset (Figure 11)
Points used in the visualisation	3.992.458	3.132.574	859.884

Table 9 The amount of points used in the visualisations in Figure 9, Figure 10 and Figure 11.

This concludes the paper. As a final note, the other attribute products of the procedure are being studied further in the hopes of unveiling a new algorithm, which can enhance the quality of the division additionally.

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Reference list

Harder et. al. Det Mangfoldige Byrum – Byrumsundersøgelse – del2. Aalborg University. Department of Architecture and Design. ISSN nr. 1603-6204. Series 19.

Jensen et. al. Identification and cleansing of scatter in GPS-based surveys in urban environments. Aalborg University. Department of Architecture and Design. ISSN nr. 1603-6204. Series 27.

Bro, Peter (2010). Cleansing GPS-data from travel surveys in urban environments, Workingpaper//preprint, AOD (Inst. 20), Aalborg University.

Anders Sorgenfri Jensen, Peter Bro, Henrik Harder, Jakob Hjorth Hansen, Nerius Tradisauskas
Diverse Urban Spaces, Aalborg University

Page 18