Capturing human activity spaces

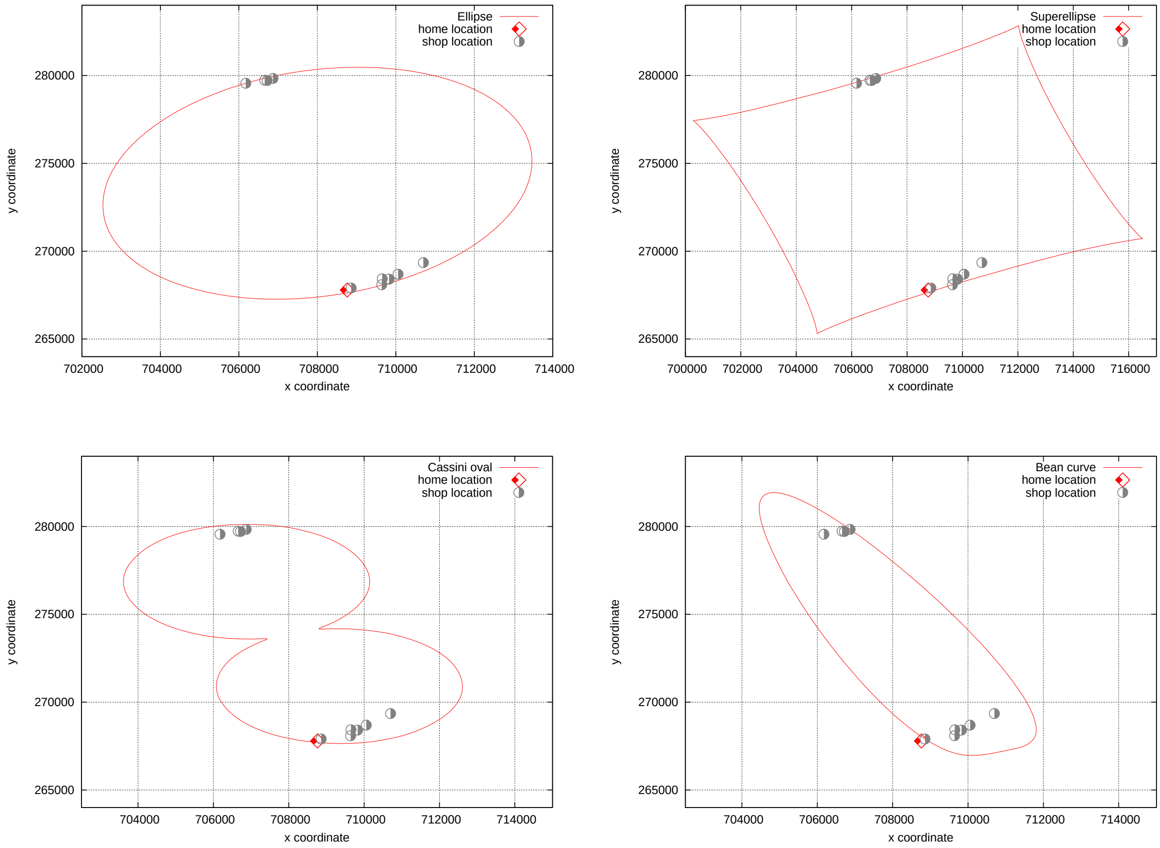
New geometries

Author(s):
Rai, Rohit K.; Rieser, Marcel; Balmer, Michael; Vaze, Vikrant S.; Schönfelder, Stefan; Axhausen, Kay W.

Publication date:
2006

Permanent link:
https://doi.org/10.3929/ethz-a-005237115

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 378



Capturing human activity spaces: New geometries

R. K. Rai
M. Balmer
M. Rieser
V. S. Vaze
S. Schönfelder
K.W. Axhausen

6000 words
1750 7 Figures / Tables

7750 words

Note: TRB paper format will be included during review process.

August 2006

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Capturing human activity spaces: New geometries

R.K. Rai
IIT Guwahati
Guwahati 781039
Assam, India

Telephone: +919864202353
rkrai@iitg.ernet.in

M. Balmer
IVT
ETH Zürich
CH-8093 Zürich

Telephone: +41-44-633-2780
Telefax: +41-44-633-1057
balmer@ivt.baug.ethz.ch

M. Rieser
VSP
TU-Berlin
D-10587 Berlin

Telephone: +49-30-314-25-258
Telefax: +41-49-30-314-26-269
rieser@vsp.tu-berlin.de

V.S. Vaze
IIT Bombay
Powai
400057 Mumbai

vikrantv@mit.edu

Stefan Schönfelder
TRAFICO Verkehrsplanung
Fillgradergasse 6/2
A-1060 Wien

Telephone: +43-1-586-4181
Telefax: +43-1-586-41-8110
stefan.schoenfelder@trafico.at

K.W. Axhausen
IVT
ETH Zürich
CH-8093 Zürich

Telephone: +41-44-633-3943
Telefax: +41-44-633-1057
axhausen@ivt.baug.ethz.ch

July 2006

Abstract

Activity space, defined as “the local areas within which people move or travel during the course of their daily activities”, is a measure of individual’s spatial behaviour which captures individual and environmental differences and offers an alternative approach to studying the spatial reach of travellers. The shape and area of the activity space is a product of how it is conceptualized and measured. The paper enlarges the set of geometries, which can be used to describe the activity space. It tests four parametric geometries (ellipse, superellipse, Cassini oval and bean curve), which are identified as those capturing a specific share of all locations visited, e.g. 95%, while minimising the area covered. They are estimated for a number of long-duration datasets while distinguishing between trip purposes.

1 Introduction

Before Geographic Information Systems (GIS) were introduced, Euclidean measures like the standard deviational ellipse (SDE), and methods making use of place-based proxies for household locations (such as zip code centroid) were used to approximate the activity space of travellers, understood here as the area which they have visited in person (See Schönfelder, 2006 for a wider review) . However, most previous studies were restricted to samples of one or two day diaries because of the expense involved in data collection, especially of the geo-coding of the addresses reported. With the recent advances in GIS technology and through the increasing availability of spatially referenced data, activity space has become a more attractive tool for studying spatial behaviour. These same technological advances enable researchers to develop new measures of activity space that improve on the precision of the standard deviational ellipse and represent and analyze actual travel behaviour in a better way.

The micro-geographical concept of activity space aims to capture the structures of the observed locational choices of the individual traveller. It is implemented, measured, with a two dimensional form (geometry) covering in a to be defined manner the places frequented by an individual over a period of time. Previously, the standard deviation ellipse (SDE), the two-dimensional generalisation of the confidence interval, was a favourite method, as it easy to calculate (See Schönfelder, 2006 for a review; Schönfelder and Axhausen, 2003 for further approaches). However, the SDE imposes a specific geometric form, which might or might not reflect the underlying behaviour or urban form. In addition, it captures the underlying variance and therefore suggests rather large areas. Furthermore, if the SDE is calculated with respect to the home location, it imposes a symmetry around the home, even if one half of the area covered is free of locations visited. Alternatives avoiding these issues would therefore be desirable.

The alternatives suggested by Schönfelder and Axhausen, 2003: the shortest-path-network or kernel-density derived measures are appealing but have computational drawbacks and, at least, the shortest path networks, substantial additional data requirements in the form of a complete navigation network for the study area. The kernel density derived measures correlate very highly with the number of observed trips and add little new information. They are, nevertheless, a good semi-parametric way to map the activity locations. If one accepts a particular geometry and a unique criterion to determine the parameters of this geometry a priori, then any geometric form could be used to capture the observed destinations: circle, square, triangle, etc. In this paper four geometries will be tested each reflecting a particular hypothesis about the form of a human activity space (see Vaze, *et al.*, 2005 for a first implementation):

- Ellipse, which can capture an activity space with either one or two clusters of locations visited combined with a range of other locations outside these clusters. (One cluster can be captured as the ellipse can collapse to a circle).
- Cassini oval, which captures again two clusters, but without intermediate locations between the clusters.
- Bean curve, which can accommodate three clusters, but is more flexible then a triangle.
- Superellipse, which includes the circle and ellipse, but can address a situation with four clusters by including a “caro” – like form.

This paper presents discusses a method to fit these geometries to observed activity locations and applies to five long-duration diaries or GPS observations. It provides in this way improved estimates of the size of the human activity space, but also for the first time into its structure. After a brief literature review, covering work from biology, geography and planning, the paper will present the algorithm implemented. The final chapter describes the data sets and the estimation results. An outlook to further work concludes.

2 Brief review

One of the first aggregate approaches to estimate people’s range of movement and contact is Hägerstrand’s Mean Information Field (MIF) (Hägerstrand, 1953). As Hägerstrand and his colleagues could not use longitudinal movement information which would have fulfilled the research requirements, they used local migration data to test the model. The concept was applied to other data sources and in different contexts later, interestingly also to one of the first longitudinal travel data sets ever, the Cedar Rapids movement study data (Garrison, *et al.*, 1959; Marble and Nystuen, 1963).

Lnych’s work (1960) focuses on the assumption that the perception of space is a highly subjective process – in contrast to the generalized representation of space in cartography. Based on the interest in the relationship between the structures as well as quality of architecture and human perception, Lynch found out that the mental maps of individuals, i.e. the image which human beings develop about their (travel) environment, are

- more or less biased
- are simplifications of the real world
- group-specific and
- composed of about five basic elements which have different meanings for the structure of urban space in different cities (paths, border lines, areas, foci and landmarks).

Mental maps mainly act as an individualised cognitive support for spatial ordering and orientation. Mental maps and their formation may methodologically be captured only indirectly – Lynch used memory protocols and - as a main approach - map sketches of test persons.

Inspired by Hägerstrand’s space-time paths (Hägerstrand, 1970), Lenntorp (1976) developed the concept of space-time prisms. He operationalised Hägerstrand’s ideas towards a measure of individual accessibility based on the notion of a person’s reach. Space-time prisms define the possible locations for a space-time path with obligatory activities such as work fixing the shape of the prism by predefining the person’s location.

Finally, the activity space concept – which was developed in parallel with several of the approaches presented above to describe individual perception, knowledge and actual usage of space in the 1960s and 1970s (see Golledge and Stimson, 1997 for a discussion) – aims to represent the space which contains the places frequented by an individual over a period of time. Activity spaces are (geometric) indicators of the observed or realised daily travel patterns (see also Axhausen, 2002). This is stressed here as the related concepts such as action space, perceptual space, mental maps or space-time prisms mainly describe the individual potentials of travel.

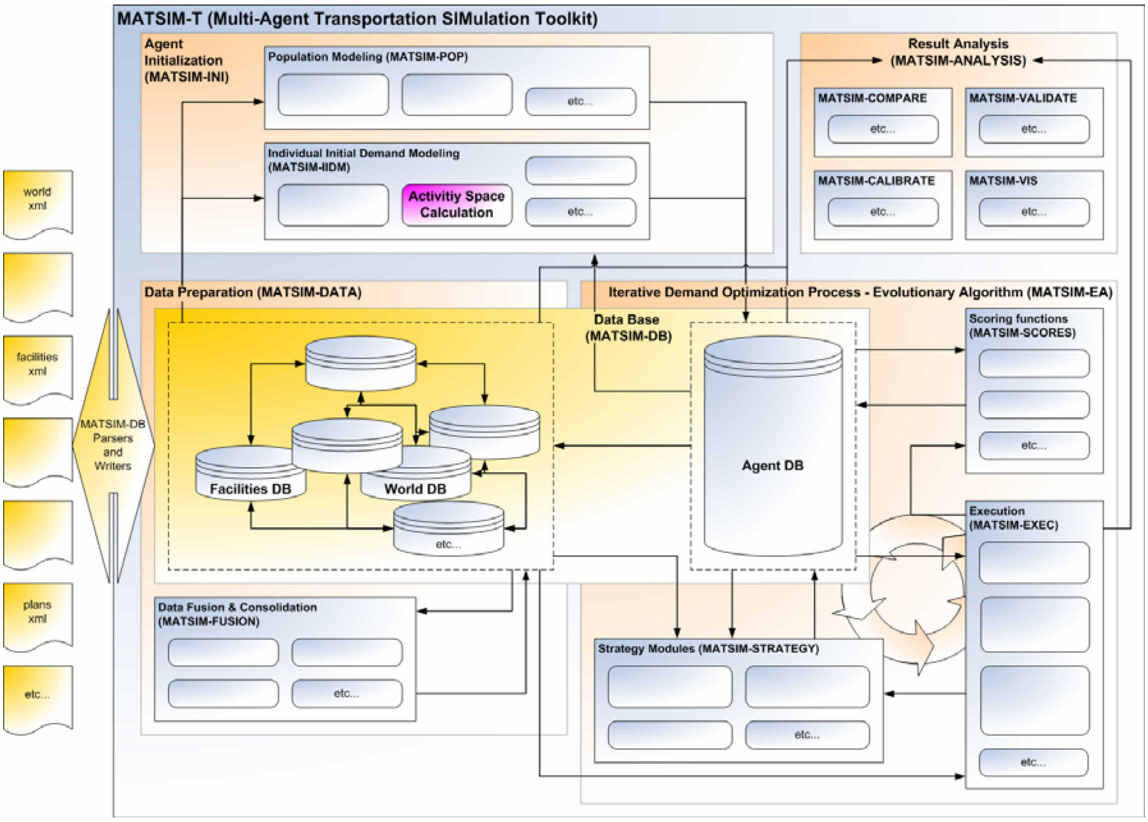
Activity spaces underlie fundamental geographical principles such as distance decay and directional bias which implies that the probability of (regular) contact with a location usually decreases with its distance from the peg(s) of daily life (i.e. in particular home) and the deviation from the main orientation / direction of daily travel. The latter refers to preferences for a particular place over other places of equal/similar distance due to some perceived quality of the preferred place (Golledge and Stimson, 1997).

3 The algorithm

The algorithm is implemented as an element of the (MATSIM-T) multi-agent transportation simulation toolkit (see Balmer, Axhausen and Nagel, 2006). The toolkit is based on a well defined database describing a given scenario which consists of spatial data, transportation networks, survey information and detailed descriptions of each individual active in the scenario (see Figure 1). The base functionality of each optional element is to read the defined XML (eXtensible Markup Language) data, store them in an appropriate data structure and write it again in an enriched, reduced, or even unchanged form in the XML data format. The person data structure is based on the MATSIM-T daily schedule DTD (in MATSIM-T called a plans-DTD) and is used as a working file to enrich persons description, their personal knowledge of the scenario and what they plan to do. In the minimum version the file holds only the identity number (id) of all persons modelled. But, obviously, it is possible to add a large amount of information about each person to the file, such as age, sex, car ownership, home, work and other locations visited by an individual etc. One part of the data structure actually defines their personal activity spaces. These data points will be calculated by the algorithm presented in this paper. The internal data structure of the plans package provides exactly the same flexibility as the XML file format. Therefore, it is possible to sequentially add additional schedule details to a given incomplete MATSIM plans file.

Algorithms – like the one described in this paper – can be added to each package to verify, manipulate, add, or delete data items according to the purpose of the algorithm. Since different algorithms have to be used or implemented for each new scenario, it is critical that the algorithms are clearly separated from the data structure. They should also be easily exchangeable.

Figure 1 Schematic overview of MATSIM-T



Adapted from Balmer (forthcoming)

3.1 Simplex optimization technique

The Nelder-Mead simplex optimization technique (Nelder and Mead, 1965) is an algorithm for finding local minimum of a function of several variables. For two variables, a simplex is a triangle, starting with three 2-dimensional points. These three points correspond to the three vertices of a triangle and constitute the first simplex. The method is a pattern search that compares function values at the three vertices of the triangle. The worst vertex, where $f(x,y)$ is largest, is rejected and replaced with a new vertex. A new triangle is formed and the search is continued (see Figure 2). The process generates a sequence of triangles, for which the function values at the vertices get smaller and smaller. The size of the triangles is reduced and the coordinates of the minimum point are found. Similarly, for a 3-dimensional space, four initial observations are required, thus defining a tetrahedral body. In general, for a function of n variables, the algorithm maintains a set of $n+1$ points forming the vertices of a simplex in n-dimensional space. This simplex is successively updated at each iteration by discarding the vertex having the highest function value and replacing it with a new vertex having a lower function value. Here is a brief outline of the steps involved in 2-dimensional simplex optimization, following Mathews and Fink (2004).

3.1.1 Initial triangle BGW

Let the function to be minimized be $f(x, y)$. Since it is a two variable function we are dealing with 2-dimensional parameter space, and hence require three initial observations to serve as the three vertices of the triangle constituting the first simplex. Let these vertices be given by $V_k = (x_k, y_k),\ k = 1, 2, 3$. The function $f(x, y)$ is then evaluated at each of the three points: $z_k = f(x_k, y_k)$ for $k = 1,2,3$. The points are then sorted according to the calculated response as best, next best and the worst. We use the notation

$$B = (x_{1,} y_1), G = (x_2, y_2), W = (x_3, y_3)$$

to help remember that B is the best vertex, G is next to best, and W is the worst (see Figure 2a).

3.1.2 Midpoint of the good side

Next step involves calculation of the mid point of the line segment joining B and G. It is found by averaging the coordinates as

$$M = \frac{(B+G)}{2} = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right)$$

3.1.3 Reflection using the point R

The function $f(x, y)$ decreases as we move along the side of the triangle from W to either of the points B and G implying that function takes on smaller values at points lying away from W on the opposite side of the line between B and G. A test point R is chosen by “reflecting” the triangle through the side BG. In order to determine R, midpoint of the side BG is calculated first. Then a line segment is drawn from W to M. Let its length be d. This line segment is extended a distance d through M to locate the point R. The vector formula of R is

$$R = M + (M - W) = 2M - W$$

3.1.4 Expansion using the point E

In case the value of the function calculated at R is smaller than that at W, it signifies that we have moved in the correct direction towards the minimum. However, since this is not the minimal point, we try to achieve a better approximation of it by extending the line segment through M and R to the point E, (see Figure 2b) forming an extended triangle BGE. The point E is found in the same manner as R by moving an additional distance d along the line joining M and R. If the value of the function is lower than that at R, it implies that E is a better vertex than R. The vector formula of E is given by

$$E = R + (R - M) = 2R - M$$

3.1.5 Contraction using the point C

If the initial reflection fails i.e. if the calculated value of the function at R is higher than that at R or if R is not within the accepted limits of parameters, a contraction is needed. A point C, called the contracted point is found out by calculating the midpoint of the line segment joining W and M. In this case the point C replaces W in the simplex. However, if the function value at C is not lower than that at W, the points G and W must be shrunk towards B. The point G is replaced with M and W is replaced with S, which is the midpoint of the line segment joining B with W (see Figure 2c).

Figure 2 Steps of the Nelder-Mead simplex-algorithm

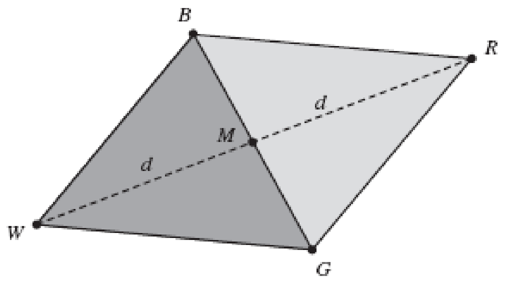


Figure 2a The triangle BGW, midpoint M and the reflected point R

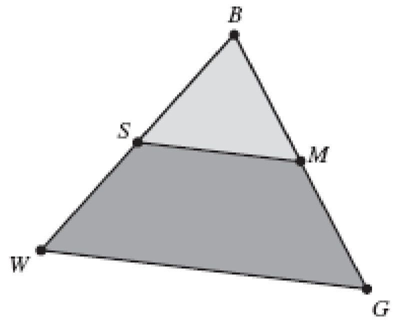


Figure 2c Shrinking of triangle towards B

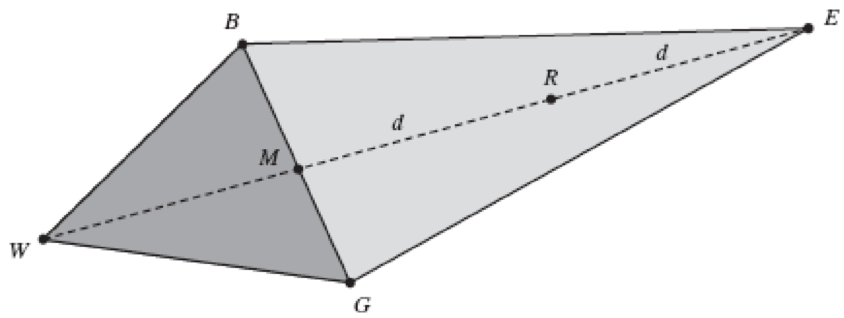


Figure 2 b The triangle BGW, point R and the extended point E

Source: Mathews and Fink (2004)

Therefore, at the end of every iteration step, we generate a new simplex with a set of three points ($n+1$ points for n dimensional simplex optimization). The process is repeated until we approach the optimum value or till the improvement of the response is insignificant between successive iterations.

3.2 Using simplex optimization for calculating activity spaces

The above described simplex optimization algorithm can be used to find activity spaces of a specified shape such that the captured area will be minimized. Furthermore we want to predefine a certain percentage of coverage of the given geo-coded locations which the resulting activity space should cover. It is therefore crucial to define which parameters are preset by the user, which ones define the parameter space in which the simplex algorithm optimizes, and—last but not least—we also need to define the objective function which has to be minimized.

The predefined input parameters of the user are:

- The type of the activity space which has to be minimized (ellipse, superellipse, Cassini oval or bean curve).
- The location type which are considered for calculating the activity space (in here we are using 'home', 'work', 'education', 'shop', 'leisure' and also the type 'all' which defines the activity space for considering all given locations).
- The minimum coverage ($cover = [0.0,1.0]$) defines the percentage of locations which has to be covered by the resulting activity space.
- The step size for angle θ, which defines the rotation of the shape based on the Euclidean coordinate system. The smaller the step size is, the more precise the resulting activity space fits the global minimum.

Note, that—in principal—the angle θ could be part of the simplex parameter space. But we already found out, that the search space including θ leads us to problems in the optimization technique. In other words, the solution space is distorted towards the direction of θ which leads to a large performance loss of the simplex algorithm.

The objective function $f(x_1, x_2, \ldots, x_n)$ calculates the captured area of the given activity space defined by its parameters $x_1, x_2, \ldots, x_n$. For all four activity spaces discussed in this paper we use the objective parameters $x0$ and $y0$ to define the centre of the activity space in the given Euclidean coordinate system. Also the horizontal and vertical extensions (a and b) are part of the search space. But since we are already set the coverage $cover$, we are able to reduce the search space by one dimension: Instead of calculating $f(x0, y0, a, b, \ldots, x_n)$, we reduce the search space such that it includes only activity spaces which covers the preset $cover$ value.

Therefore, we substitute a and b by the ratio $ratio = b/a$ and let the objective function calculate a and b, such that it fulfils $ratio = b/a$ and covers the defined amount of locations (preset by the $cover$ parameter). As a result the search space of the simplex algorithm is defined by the coordinates $x0$ and $y0$, the ratio $ratio = b/a$ and—dependent on the given shape of the activity space—additional parameters.

The objective function $f(x0, y0, a, b, \ldots, x_n)$ is therefore replaced by $f_{cover}(x0, y0, ratio, \ldots, x_n)$. In the objective function, the calculation of the actual extends of a and b will be done via bisection method.

The detailed description of the search space of each evaluated shape is described in the next section.

3.3 Geometries implemented

As discussed above, the geometries chosen impose a parametric form on the activity space. Although these geometries are an abstract representation of travellers' activity spaces, they actually allow for construction of a more comprehensive and realistic picture of travel behaviour than has previously been provided. They provide an understandable graphical representation of a combination of elusive concepts and also several appropriate measures for quantification and comparability to allow further analysis of activity spaces. This section deals presents the mathematical definition of the geometries (see Figure 3 for their graphical representations).

3.3.1 Ellipse

Ellipse is defined as the locus of points P such that the sum of the distances from P to two fixed points F_1, F_2 (called foci) is constant. That is, $\text{distance}[P, F_1] + \text{distance}[P, F_2] = 2a$, where a is a positive constant. An ellipse centred at the origin of an $x-y$ coordinate system with its major axis along the x-axis is defined by the equation of the elliptical object.

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

The same ellipse is also represented by the parametric equations:

$$x = a\cos t; \quad y = b\sin t\ ,$$

where t belongs to $[0,2\Pi]$

If the ellipse is not centred at the origin of x-y coordinate system but has its major axis along the x-axis, it may be defined as

$$\left(\frac{x - x0}{a}\right)^2 + \left(\frac{y - y0}{b}\right)^2 = 1\,,$$

where $(x0, y0)$ is the centre.

The parametric form of an ellipse centred at $(x0, y0)$ and rotated through an angle θ is given by

$$x = a\cos t\cos\theta - b\sin t\sin\theta + x0, \quad y = a\cos t\sin\theta + b\sin t\cos\theta + y0$$

The areas of an ellipse is given by

$$A = \Pi \times a \times b$$

In order to find the ellipse covering 95% of the activity locations, four parameters were fitted:

1. X coordinate of the centre of the ellipse ($x0$).

2. Y coordinate of the centre of the ellipse ($y0$).

3. The orientation of the major axis of the ellipse (θ).

4. The ratio ($ratio$) of the length of the semi-minor axis (b) to the length of the semi-major axis (a) of the ellipse.

3.3.2 Superellipse

A superellipse (Lamé curve), centred at origin, is defined in the Cartesian coordinate system as the set of points satisfying the equation

$$\left|\frac{x}{a}\right|^r + \left|\frac{y}{b}\right|^r = 1,$$

where $r \rangle 0$ and a and b are the radii of the oval shape. The case $r = 2$ yields an ordinary ellipse; r values below 2 result in hyper-ellipses with pointed corners in the x and y directions resembling crosses, r values greater than 2 yield hyper-ellipses which increasingly resemble rectangles.

It may be described parametrically as

$$x = a\cos^{2/r} t, \quad y = b\sin^{2/r} t$$

A superellipse centred at $(x0, y0)$ is given by

$$\left|\frac{x - x0}{a}\right|^r + \left|\frac{y - y0}{b}\right|^r = 1,$$

with parametric equation

$$x = a\cos^{2/r} t + x0, \quad y = b\sin^{2/r} t + y0$$

A superellipse with centre at $(x0, y0)$ and having an orientation of θ with the x-axis is defined parametrically as

$$x = a\cos^{2/r} t\cos\theta - b\sin^{2/r} t\sin\theta + x0, \quad y = a\cos^{2/r} t\sin\theta + b\sin^{2/r} t\cos\theta + y0$$

The area of the superellipse is given by

$$A = \frac{4^{1-1/r} ab\sqrt{\Pi} * \Gamma(1 + 1/r)}{\Gamma(1/2 + 1/r)},$$

where Γ is the gamma function.

To construct optimum superellipse, the following parameters were calculated:

1. X coordinate of the centre of the superellipse ($x0$).

2. Y coordinate of the centre of the superellipse ($y0$).

3. The orientation of the major axis of the superellipse (θ).

4. The ratio ($ratio$) of the length of the semi-minor axis (b) to the length of the semi-major axis (a) of the superellipse.

5. The exponent r .

The exponent r defines if a superellipse looks like a diamond ($0 < r < 1$), like a rectangle ($r = 1$) or more like an ellipse ($r > 1$). In this paper we are interested into diamond shapes of the superellipse, therefore we reduce the search space of the simplex algorithm to values of r between zero and one.

3.3.3 Cassini oval

The Cassini ovals are defined as a locus of a point such that the product of its distances from two fixed points a distance $2a$ apart is a constant b^2. If the long axis of symmetry of the curve centred at origin, is parallel to the *x*-axis, it can be expressed as

$$\left((x-a)^2 + y^2\right)\left((x+a)^2 + y^2\right) = b^4$$

The parametric form of the above curve is:

$$x = a\cos t\sqrt{\cos 2t + \sqrt{\left(\frac{b}{a}\right)^4 - \left(\frac{\sin 2t}{a}\right)^2}}, \quad y = a\sin t\sqrt{\cos 2t + \sqrt{\left(\frac{b}{a}\right)^4 - \left(\frac{\sin 2t}{a}\right)^2}}$$

The parametric equation of a cassini at an angle θ and centred at $(x0, y0)$ is:

$$x = a\cos(\theta + t)\sqrt{\cos 2t + \sqrt{\left(\frac{b}{a}\right)^4 - \left(\frac{\sin 2t}{a}\right)^2}} + x0;\ y = a\sin(\theta + t)\sqrt{\cos 2t + \sqrt{\left(\frac{b}{a}\right)^4 - \left(\frac{\sin 2t}{a}\right)^2}} + y0,$$

The area of the curve is given by

$$A = \frac{1}{2} r^2 d\theta = \int_{-\Pi/4}^{\Pi/4} a^2 \left[\cos(2\theta) \pm \sqrt{\left(\frac{b}{a}\right)^4 - \sin^2(2\theta)} \right] d\theta$$

The shape of the oval depends on the ratio b/a. When b/a is greater than 1, the locus is a single, connected loop. When b/a is less than 1, the locus is comprises two separated parts as shown in Figure 3. When b/a is equal to 1, the locus is a lemniscate. Since, the situation b/a less than 1 is not relevant to our purpose; it was excluded by constraining b/a to be greater than 1 throughout the optimization process.

For the Cassini oval, the optimization process fitted the following four parameters:

1. X coordinate of the centre of the cassini oval ($x0$).

2. Y coordinate of the centre of the cassini oval ($y0$)

3. The orientation of the major axis of the cassini oval (θ).

4. The ratio ($ratio$) of the length of the semi-minor axis (b) to the length of the semi-major axis (a) of the cassini oval.

3.3.4 Bean curve

The standard unit bean curve, situated in the 1st and the 4th quadrant, with origin at its one end and having the horizontal axis of unit length oriented along $x-$axis, is given by the equation

$$x^4 + x^2 y^2 + y^4 = x\left(x^2 + y^2\right)$$

The parametric form for the same is:

$$x = \sin^2\theta, \quad y = \sin\theta\sqrt{\frac{\cos^2\theta + \cos\theta\sqrt{1+3\sin^2\theta}}{2}}$$

The area of the bean curve is given by

$$A = \sqrt{2}\int_0^1 \sqrt{x\left(1 - x + \sqrt{1+(2-3x)x}\right)}dx \cdot a \cdot b \simeq 1.058049 * a * b$$

The optimization is performed in a 4-parameter space consisting of the following:

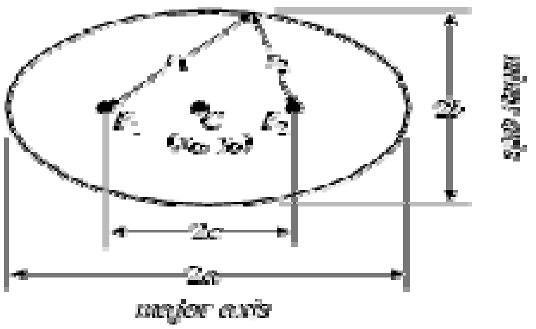
1. X coordinate of the centre of the bean ($x0$).

2. Y coordinate of the centre of the bean ($y0$).

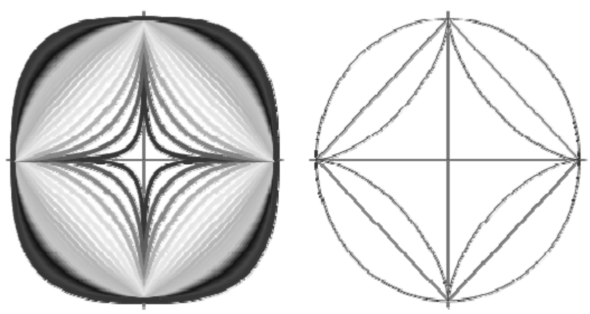
3. The orientation of the major axis of the bean (θ).

4. The ratio b/a, where a and b are the multiples by which the standard unit bean curve is stretched in the horizontal and vertical directions respectively.

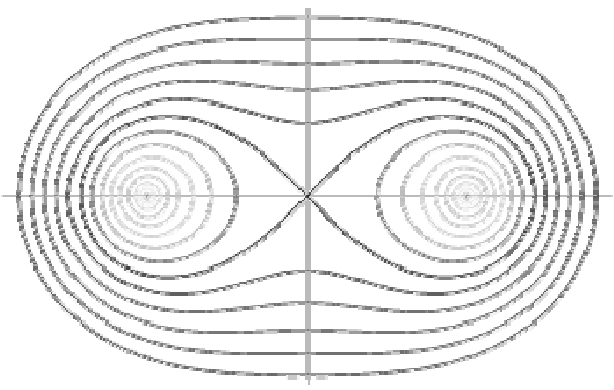
Figure 3 Geometries tested



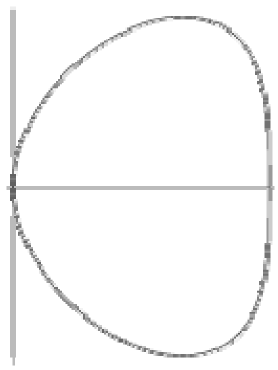
Ellipse



Superellipse



Cassini oval



Bean curve

Source: http://mathworld.wolfram.com

4 Results and discussion

4.1 Data sources

In the past, transport planners collected short duration, generally one-day diaries, because of the cost and effort required for longer duration studies. However, with the availability of GPS-based tracking system and the increased interest in studying travel behaviour, a number of long duration surveys and observational studies have become available.

The data sets employed here are the Uppsala 5-week diary (Hanson and Burnett, 1982), Mobidrive 6-week diary (Axhausen *et al*, 2002), ISA Rättfart GPS observational study (Schönfelder *et al* 2002; Schönfelder and Samaga, 2003), SVI Stabilität 6-week diary (Axhausen *et al*., 2004) and the AKTA GPS observational study (Nielsen and Jovicic, 2003).

The Uppsala survey was conducted in the city of Uppsala, located approximately 70 km northwest of Stockholm, in the year 1971. A random sample of 20% of the population was drawn. The final sample size was 278 households with 488 persons of which 92 households were chosen for further analysis. The addresses of all trips were geocoded by hand.

The Mobidrive survey, conducted in the German cities of Halle/Saale and Karlsruhe in 1999, involved a total of 317 persons over 6 years in 139 households. The trip destination addresses of all main study trips were geo-coded with the geo-coding being positive for more than 98%

of the trips. For the City of Karlsruhe and the City of Halle, small street blocks were used as the basis for the geo-coding of the street addresses. However, outside the urban area, the addresses were geo-coded only on the basis of the centroid of the municipality.

The ISA Rättfart GPS study, carried out in the Middle-Swedish town of Borlänge, consists of fully automatically collected movement information for vehicles for up to two years. The vehicles were equipped with on-board data collection system consisting of a GPS receiver, a data storage device running a GIS for mapping all movements and a mobile power supply. The study was conducted from 1999 to 2002 with more than 200 private and commercial cars equipped for periods of up to two years each.

The SVI Stabilität survey was performed in the canton of Thurgau in 2003, covering a six week reporting period with 99 households and 230 persons. Nearly all destination addresses and household locations could be geo-coded with high precision.

In the AKTA study, carried out in the greater Copenhagen region, approximately 400 cars were equipped with a GPS-based device during a period of two times 8-10 weeks in 2001/2002. Vehicle movement data was collected each second. A telephone based before-and-after survey which consisted of attitude questions accompanied the process of GPS monitoring. The available sample includes 50 vehicles/persons with 44 to 135 reported days and 125 to 1044 reported trips each. The trip was defined by the first satellite signal received and ends with the engine switch off. The visited locations were identified by a simple clustering technique which grouped adjacent trip ends into clusters using the SAS Fastclus procedure (Anderberg, 1973).

4.2 Setup

For all the above given surveys we will calculate the 95% coverage activity spaces ($cover = 0.95$). It has to be noted, that it is typically not possible to obtain exactly the given coverage because the number of locations excluded are discrete numbers. Therefore, the algorithm will calculate the lowest possible coverage above 0.95.

Each location of the above studies includes the frequency how often a person went to that location. We are using this as the weight of the location by simply multiply it by the frequency. With it, locations which are visited quite often will typically be part of the calculated activity space with coverage 0.95.

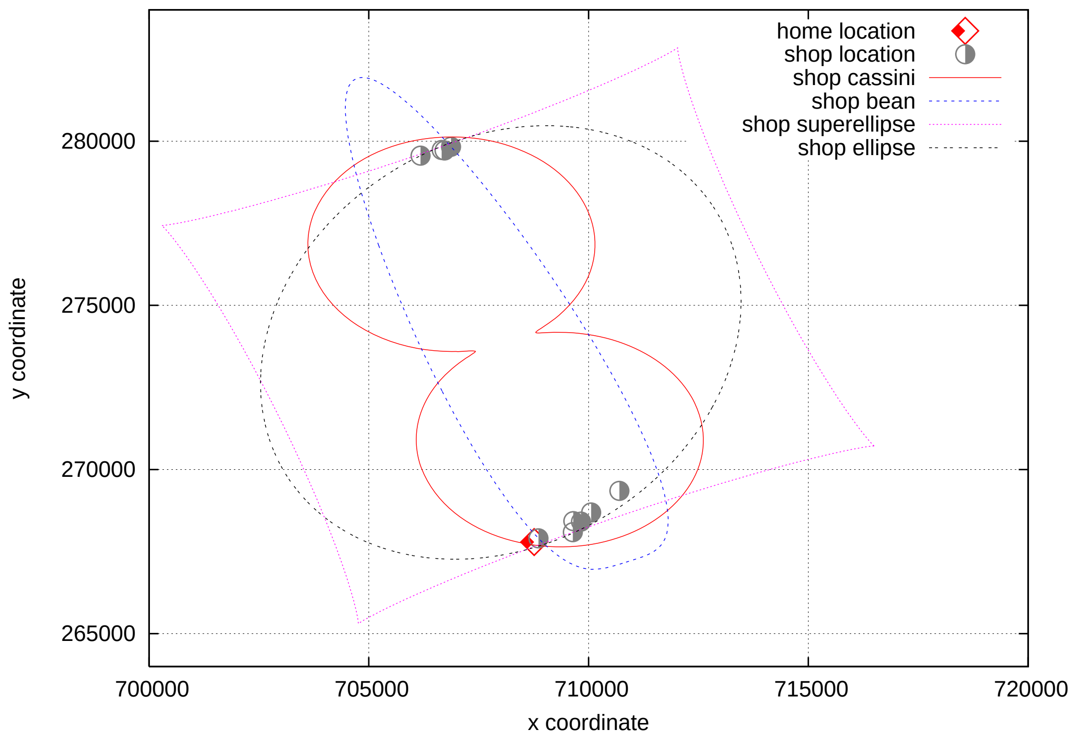
We assigned a orientation step size of θ by $\theta_{step} = 22.5°$. This is a good trade off between computation time and accuracy of the resulting activity space.

4.3 Results

Before discussing the results for the whole samples, a set of example results will be discussed and mapped. Figure 4 depicts the different optimal geometries covering 95% of the shopping locations visited by a respondent drawn from the “SVI Stabilität” study. The person reported 218 trips, sorted by the activity types:

- 44 leisure activities at 24 unique leisure locations
- 77 work activities at 11 unique work locations
- 84 home activities at 1 unique home locations
- 12 shop activities at 10 unique shop locations
- 1 education activities at 1 unique education locations

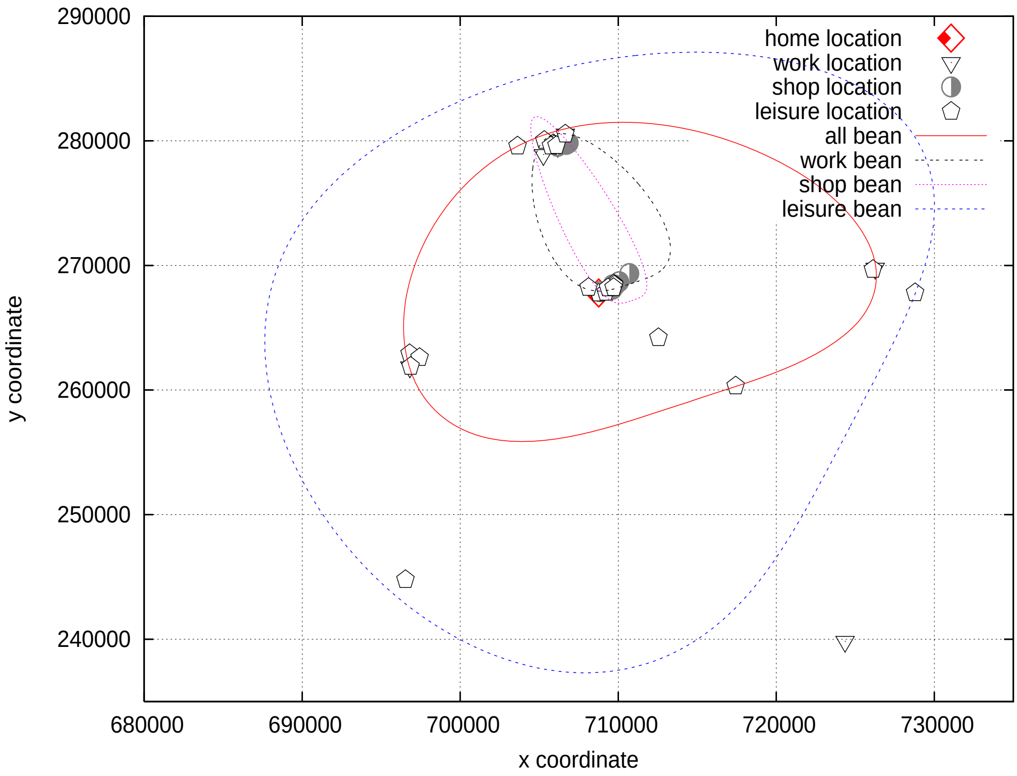
Figure 4 Example optimal geometries for the shopping locations plus home



The point to be noted here is that the superellipse produced is not the "best" optimum one that is possible. This is due to the limited number of angles which were tested during the optimisation (see above). A smaller step size for angles could be adopted for higher accuracy but at the cost of increased computation time.

Figure 5 shows the optimal bean-curves for all locations, as well as for the four activity purpose subsets. The bean curve covering 95% of all activity locations is smaller than that for leisure activities, as there are peripheral and rarely visited leisure locations.

Figure 5 Example 95% coverage activity spaces by weighted purpose (bean curves)



The optimal parameters were fitted for all geometries proposed and all five datasets (Table 1). In addition, the best, i.e. smallest, of the four geometries was determined for each respondent. All distributions are highly left skewed with substantially smaller medians than means. The large coefficients of variation underline this variance. While the mean and median of the best-of geometry is much smaller then any of the other distributions, as expected. This highlights that no one geometry is appropriate for all respondents. The Cassini fits in about two of three cases, but not always. It indicates that the Cassini oval best reflects the typical pattern of two dominant clusters of activity (See Schönfelder, 2006), mostly home, work and education, but also frequently a leisure oriented cluster. In less then 10% of the cases, the bean curve capture a pattern with three clusters.

Table 1 Optimal geometries for all locations: Summary of results

Uppsala diary	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	115.71	197.80	145.40	200.55	92.93
Median	9.04	16.10	13.70	18.40	8.96
Standard deviation	354.80	652.46	508.67	674.88	326.45
Frequency of being the best solution	113	7	10	14	144

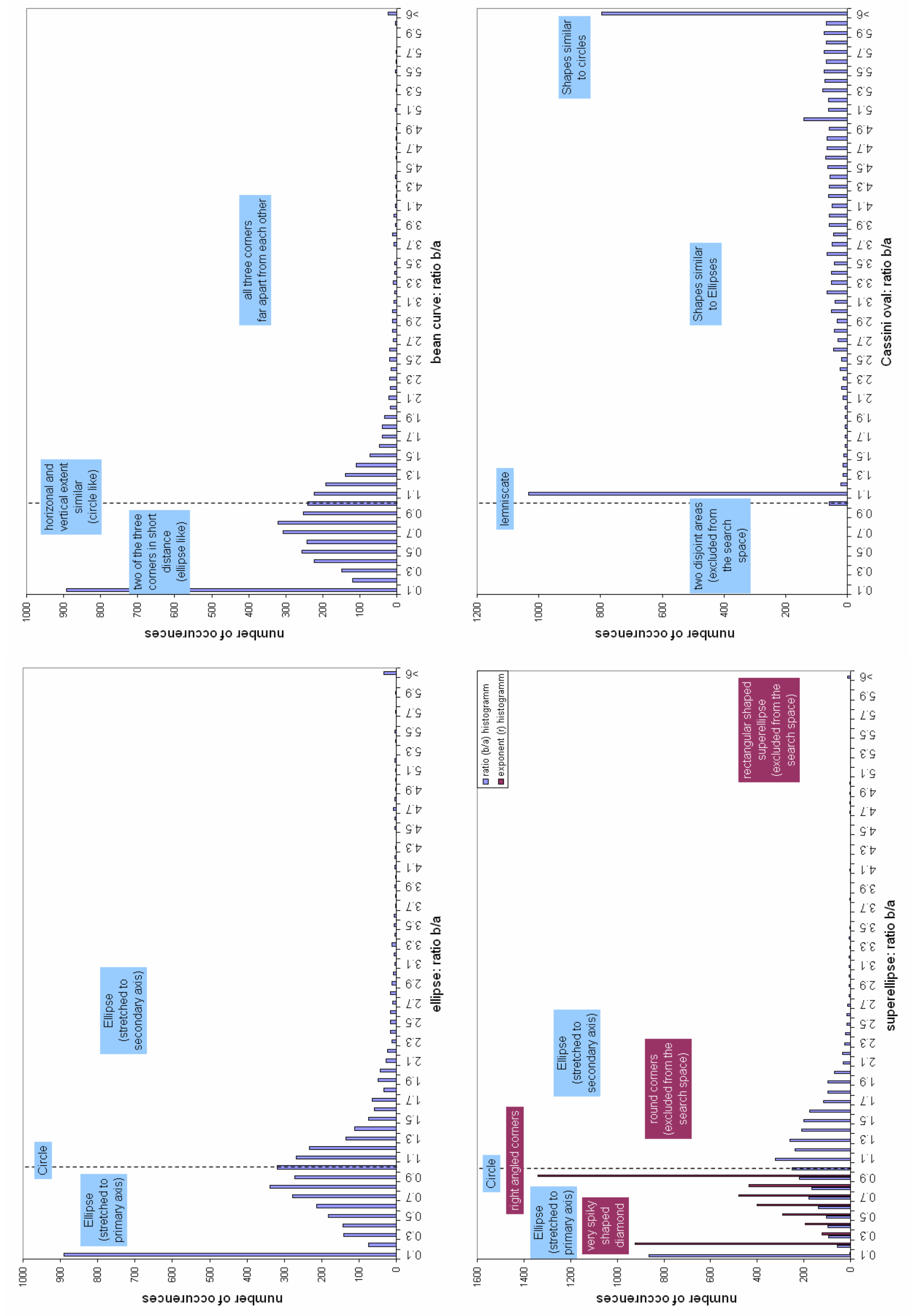
Mobidrive diary	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	384.74	629.12	597.88	558.71	333.63
Median	42.30	71.50	66.70	76.90	37.60
Standard deviation	1909.28	3175.02	3166.75	2846.46	1665.40
Frequency of being the best solution	257	16	20	24	317

ISA Rättfart GPS study, Borlänge	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	102.10	141.20	165.43	147.98	88.79
Median	65.60	110.00	102.00	112.00	65.60
Standard deviation	128.44	110.23	214.70	126.55	92.48
Frequency of being the best solution	53	1	5	7	66

SVI Stabilität survey, Thurgau	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	1845.99	2819.77	2509.13	3419.68	1724.84
Median	347	598	460	637	319
Standard deviation	3893.65	6123.99	8087.50	3667.79	3219.41
Frequency of being the best solution	177	14	26	13	230

The AKTA study, Copenhagen	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	3322.98	4111.08	5013.48	4869.48	2603.34
Median	246.00	426.00	366.00	491.00	241.00
Standard deviation	7520.15	10729.63	11684.94	13688.16	6314.43
Frequency of being the best solution	146	26	17	12	201

Figure 6 Occurrence distributions of ratio (b/a) for the four shapes



If we compare the number of occurrences of a specific ratio of one of the given activity spaces, as shown in Figure 6, there are several aspects to mention:

1. In the ellipse, superellipse and the bean curve histograms there area high amount of ratio (b/a) almost equal to zero. There are artefacts created by calculating activity spaces for one or two geo-coded locations. Since those activity spaces do not have any use, we do not consider this part in the analysis.

2. The ratio histogram of the ellipse shows more or less a normal distribution with the mean of ratio b/a equals to 1. In other words, many resulting ellipse activity spaces capture an area similar to a circle. With the information given by Table 1 we already know, that there are many exceptions in which the ellipse type does not cover an appropriate area while other shapes minimizes the space better.

3. Similar can be said about the superellipse. While the ratio is normal distributed around b/a = 1, the exponent (r) are either very small (which forms a very spiky shaped diamond) or almost 1 (the shape of a “caro”). Since the analysis in Table 1 showed that the superellipse does not capture an appropriate activity space compared to the other shapes, it is no improvement of the ellipse shape.

4. On the other hand, a very promising output is delivered by the Cassini oval. The majority parts of the shapes are lemniscates. The others tent to form a circle. This indicates that the Cassini oval nicely captures location sets which are concentrated at two spatial regions. The ellipse and the superellipse fail in such cases. It has to be noted, that the Cassini oval does—per definition—consume a low amount of space compared to the ellipse. Therefore, it is not that surprising that the Cassini oval has a high frequency of being the best solution.

5. Last but not least, the bean curve also shows a promising histogram. Since its shape is the most flexible of the four tested ones, we can capture locations sets with two groups (b/a < 1) as well as set with three groups (b/a > 1). Also circle like activity spaces can be captured (b/a ~ 1). A two-group location set are better captured by the Cassini oval, but the bean curve is better trimmed as the ellipse.

5 Summary and Outlook

In this paper we presented a generic activity space calculation module embedded in the MATSIM-T project. It has the flexibility to produce an appropriate activity space representation for a given location set. The module can be varied by the type of the shape of the activity space, the location type which has to be covered by the activity space and the percentage of the locations which the activity space should cover. The module can be used at any time of the initial individual demand modelling process of MATSIM-T to enrich the person data structure of the MATSIM Database.

But the use of the simplex algorithm to obtain the minimum activity space for a given shape fulfils the requirement only partially. It is unsatisfying to preset the orientation of the shape stepwise and then pick the best solution. Some tests by adding the orientation into the search space showed clearly, that the simplex algorithm could stuck into a local optimum. To gain more flexibility (and at the end also gain computational speedup) it is necessary to replace the simplex algorithm by a more stable optimization model which can handle more complex search spaces. Fairly good approaches are evolutionary strategies like the covariance matrix adaptation (CMA) as used in Charypar, et al. (2006).

The results in the previous chapter showed, that the bean curve and the Cassini oval produces more appropriate activity spaces as the ellipse does. It would be desirable do combine the two shape to gain the advantages of both. We may want to implement shapes which are less specialised and are able to define activity space much more efficiently and accurately.

Also, it would be interesting to see as to what factors influence the activity space sizes of different individuals. A detailed statistical analysis needs to be performed to see if the socio-demographic variables affect the travel behaviour and—with it—their individual activity space.

The empirical results reported in this paper suggest that the activity space concept has the potential of becoming a more widely used tool in studying spatial access, transport policy and planning and can be used to understand urban travel behaviour. It can be used to evaluate present and future urban structure and accordingly come up with solutions to satisfy the activity demand in the household’s neighbourhood, resulting in reduced travel expenses, congestion and emissions. With the growing recognition of need for long duration datasets, the future analyses will lead to improved understanding of urban travel behaviour and thus more accurate forecasts and enhanced policy analyses.

6 Acknowledgements

The Uppsala travel diary data was made available by Prof. Susan Hanson, Clark University, Worcester.

The transport psychologists from the universities of Dalarna and Uppsala made available the ISA Rättfart GPS study dataset.

The Centre for Traffic and Transport of the Technical University of Denmark kindly provided the Copenhagen AKTA data.

7 Literature

Anderberg, M.R. (1973), Cluster Analyis for Applications, Academic Press, New York.

Axhausen, K.W. (2002) A dynamic understanding of travel demand: A sketch, *Arbeitsberichte Verkehrs- und Raumplanung*, **119**, IVT, ETH, Zürich.

Axhausen, K. W., A. Zimmermann, S. Schönfelder, G. Rindsfüser and T. Haupt (2002) Observing the rhythms of daily life: A six-week travel diary, *Transportation*, **29** (2) 95-124.

Axhausen K.W., M. Löchl, R. Schlich, T. Buhl and P. Widmer (2004) Fatigue in long-duration travel diaries, Arbeitsbericht Verkehrs- und Raumplanung, **286**, IVT, ETH Zürich, Zürich.

Balmer, M., K. W. Axhausen and K. Nagel (2006) A demand generation framework for large scale micro-simulations, in TRB 85th Annual Meeting Compendium of Papers CD-ROM, Transportation Research Board, Washington, D.C.

Balmer, M. (forthcoming) Travel demand modeling for multi-agent trafficsimulations: Algorithms and systems, Ph.D. Thesis, ETH Zurich, Zurich.

Charypar, David, Axhausen, Kay W., and Nagel, Kai (2006) Implementing Activity-Based Models: Accelerating the Replanning Process of Agents Using an Evolution Strategy, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, August 2006.

Garrison, W.L., B.J.L. Berry, D.F. Marble, J.D. Nystuen and R.L. Morrill (1959) Studies of highway development and geographic change, University of Washighton, Seattle.

Golledge, R.G. and R.J. Stimson (1997) *Spatial Behavior*, The Guilford Press, New York.

Hägerstrand, T. (1953) *Innovationsforloppet ur korologisk synpunkt*, Gleerup, Lund.

Hägerstrand, T. (1970) What about people in regional science?, *Papers of the Regional Science Association*, **24** (1) 7-21.

Hanson, S. and K.O. Burnett (1982) The analysis of travel as an example of complex human behaviour in spatially-constraint situation: Definition and measurement issues, *Transportation Research A*, **16** (2) 87-102.

Kitamura R, L.P. Kostyniuk and M.J. Uyeno (1981) Basic properties of time-space paths: empirical tests. *Transportation Research Record No.794* (pp 8-19). Washington: Transportation Research Board.

Kostyniuk L.P. and R. Kitamura (1982) life cycle and household time-space paths: empirical investigation. *Transportation Research Record No.879* (pp 28-37). Washington: Transportation Research Board.

Lenntorp B (1976) *Paths in Space-Time Environments: a Time-Geographic Study of the Movement Possibilities of Individuals.* Lund Studies in Geography, Series B, Number 44

Lenntorp B (1978) A time geographic simulation model of individual activity programmes. In: Carlstein T, Parkes D & Thrift N (eds) *Human Activity and Time Geography* (ch.9). London: E.Arnold.

Lynch D. (1960) *The Image of the City*, MIT-Press, Cambridge.

Marble, D.F. and J.D. Nystuen (1963) An approach to the direct measurement of community mean information fields, *Papers and Proceedings of the Regional Science Association*, **11**, 99-109.

Mathews, J.H. and Kurtis K. Fink. Numerical Methods Using Matlab, 4th Edition, 2004, Prentice Hall, Upper Saddle River, New Jersey

MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com, as accessed on 20 May 2006

Nelder, J.A. and R. Mead (1965). A simplex method for function minimization. Computer Journal, **7**, 308-313.

Nielsen, O. A. and G. Jovicic (2003) The AKTA Road Pricing Experiment in Copenhagen, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

Schönfelder, S. (2006) Urban rythms, PhD dissertation, ETH Zürich, Zürich.

Schönfelder, S. and K.W. Axhausen (2003) Activity spaces: Measures of social exclusion ?, *Transport Policy,* **10** (4) 273-286.

Schönfelder, S., K.W. Axhausen, N. Antille und M. Bierlaire (2002) Exploring the potentials of automatically collected GPS data for travel behaviour analysis - A Swedish data source, in J. Möltgen und A. Wytzisk (Hrsg.) *GI-Technologien für Verkehr und Logistik*, IfGIprints, **13**, 155-179, Institut für Geoinformatik, Universität Münster, Münster.

Schönfelder, S. and U. Samaga (2003) Where do you want to go today? - More observations on daily mobility, paper presented the 3rd Swiss Transport Research Conference, *Arbeitsberichte Verkehrs- und Raumplanung*, **179**, Ascona, March 2003.

Vaze V.S., S. Schönfelder and K.W. Axhausen (2005) Optimization of continuous space representation for human activity spaces, *Arbeitsbericht Vekehrs- und Raumplanung*, **295**, Institut für Verkehrsplanung and Transportsysteme (IVT), ETH Zürich, Zürich