

Geometric methods for estimating representative sidewalk widths applied to Vienna's streetscape surfaces database

Tadej Brezina¹  · Anita Graser² · Ulrich Leth¹

Received: 24 May 2016 / Accepted: 27 December 2016 / Published online: 22 February 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Space, and in particular public space for movement and leisure, is a valuable and scarce resource, especially in today's growing urban centres. The distribution and absolute amount of urban space—especially the provision of sufficient pedestrian areas, such as sidewalks—is considered crucial for shaping living and mobility options as well as transport choices. Ubiquitous urban data collection and today's IT capabilities offer new possibilities for providing a relation-preserving overview and for keeping track of infrastructure changes. This paper presents three novel methods for estimating representative sidewalk widths and applies them to the official Viennese streetscape surface database. The first two methods use individual pedestrian area polygons and their geometrical representations of minimum circumscribing and maximum inscribing circles to derive a representative width of these individual surfaces. The third method utilizes aggregated pedestrian areas within the buffered street axis and results in a representative width for the corresponding road axis segment. Results are displayed as city-wide means in a 500 by 500 m grid and spatial autocorrelation based on Moran's I is studied. We also compare the results between methods as well as to previous research, existing databases and guideline requirements on sidewalk widths. Finally, we discuss possible applications of these methods for monitoring and regression analysis and suggest future methodological improvements for increased accuracy.

Keywords Streetscape · Urban surface database · Sidewalk width · Inscribed circle · Circumscribed circle

✉ Tadej Brezina
tadej.brezina@ivv.tuwien.ac.at

¹ Research Center for Transport Planning and Traffic Engineering, Institute of Transportation, Vienna University of Technology, Gußhausstraße 30/230-1, 1040 Vienna, Austria

² Business Unit Dynamic Transportation Systems, Mobility Department, Austrian Institute of Technology, TECHbase Vienna, Giefinggasse 2, 1210 Vienna, Austria

JEL Classification R14 · R41 · R52

1 Introduction

Today's IT capabilities and data collection on various urban management levels provide mobility relevant data on a comprehensive basis. In the past, random tape measurements provided anecdotal evidence of characteristic dimensions of transport infrastructure. Today's new geographic datasets enable systematic and comprehensive derivation of characteristic measurements. Besides detailed measurements for planning, the need of area-wide, quick, workable and relation-preserving overview methods may prove more important to some management needs than in-depth analysis. GIS is increasingly being used for deriving transport-related state variables, e.g. catchment areas of public transport stations (Iseki and Tingstrom 2014), the accessibility to public transport (O'Sullivan et al. 2000) or network performance (Mesbah et al. 2012). Walkability is another parameter, which has gained importance in recent quantitative analyses, for example analysing the connection between quality of urban built environment and mental health in youth (Duncan et al. 2013), active school transport (Wong et al. 2011), obesity (Duncan et al. 2012; Agampatian 2014) or the dichotomy of objective and perceived walking times (Dewulf et al. 2012). Findings suggest strong correlations between walkability and likelihood of taking a walking trip, vehicle miles travelled and obesity prevalence (Frank et al. 2007) and with moderate-to-vigorous physical activity (Sallis et al. 2009). Due to the lack of more detailed data, a lot of studies stay on a meta-level when linking walkability with indicators such as land-use mix, residential density and intersection density (Frank et al. 2005) or pedestrian route directness. These can be calculated easily using GIS software and existing data. However, even simple indicators such as mere sidewalk presence or sidewalk width—used in a few studies (Ewing et al. 2004; Lin and Chang 2010)—need to be collected at least in part manually, which is time-consuming and costly (Schneider et al. 2005; Frackelton et al. 2013). Systemic data collection on sidewalk widths is needed. Here, our paper steps in with the proposition of automated computation of representative sidewalk widths with differently grained methods according to data availability.

Our paper is organized as follows: the next section presents three methods to compute representative sidewalk width and applies them to a dataset from Vienna, Austria. Section 3 presents distribution parameters of resulting representative widths and the mapped study of autocorrelation in terms of position in the city. The second to last section discusses the plausibility of results and draws conclusions on result significance elaborately. In the final section, results are reviewed as well as method application and improvement is debated.

2 Materials and methods

2.1 The dataset

We use the SIS_F urban surface management database of the Vienna's road administration, which includes all public space surfaces within Vienna's municipal

borders classified by type, pavement, maintenance responsibility, etc. This database contains 804,958 surface polygons and includes surface types such as road lanes, stairs, tramway rights of way, street islands and cycle tracks. Surface types that cover sidewalk areas are: sidewalk (GG), driveway entrance for house access (EE) and walkway entrance for house access (HH), which we use for our analysis. With few exceptions, these three types cover all walkable surfaces between facades and road lanes. The analysis does not include pedestrian precincts (FZ) since they tend to cover the total street width from facade to facade and separated walkways (FW) and since they denominate alignments detached from the regular streetscape, e.g. paths through parks. Table 1 describes the dataset (number of surfaces per type) as well as characteristic parameters (min, max, mean) of area and circumference.

In effect, minimum areas are not zero, but smaller than 100 cm². As not only physical boundaries are shown in the SIS_F database, but also lines connecting house corners randomly delimit surfaces, such extremely small snippets exist. For example, the lower left corner of Fig. 2 shows such a triangular snippet—in some cases their size is less than the threshold given above.

A first analysis for Vienna's 23 administrative districts (Fig. 1) suggests that increasing area size and distance from the city centre cause a sub-linearly growing portion of pedestrian surfaces, with an exponent of 0.77. This sub-linear development shows that peripheral districts with more recent road designs provide less pedestrian surfaces than central areas with often century old road layouts.

2.2 Three methods

We derived and applied three different methods for estimating representative sidewalk width w_{REP} . Methods 1 and 2 calculate w_{REP} per individual surface, while Method 3 aggregates on a street axis basis:

- Method 1 uses the minimum circumscribed circle of a polygon.
- Method 2 uses the maximum inscribed circle of a polygon.
- Method 3 utilizes the total sidewalk area within a buffer around the street axis.

Table 1 Dataset parameters

Surface type	GG	EE	HH
Sample size	349,462	77,950	91,623
Area (m ²)			
Min	0.00	0.00	0.00
Mean	25.16	8.48	3.84
Max	2739.44	803.90	330.12
Circumference (m)			
Min	0.01	0.04	0.02
Mean	26.33	11.80	7.78
Max	1636.47	307.47	84.15

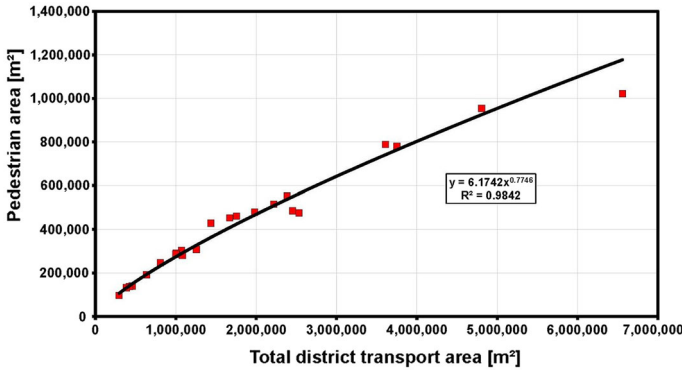


Fig. 1 Pedestrian areas (GG + EE + HH) over total district transport surface

The first two methods keep the distinction between the three surface types as given by the SIS_F database for reasons of precision and small-sized permanent changes in in situ widths. The different orientation of length versus width calls for the necessity to different methodical approaches. An inclusion of EE or HH surfaces to adjoining GG surfaces would reduce the detailed representation of constant width changes.

In Method 1, the representative sidewalk width is calculated as

$$w_{REP,EE/HH} = d_C \quad \text{and} \quad w_{REP,GG} = \frac{A_{GG}}{d_C}, \tag{1}$$

where A_{GG} is the surface area of the type GG shape and d_C is the diameter of the circumscribed circle. $w_{REP,EE/HH}$ is directly associated with d_C as EE and HH type surfaces are mostly oriented across the logical sidewalk width. Since GG surfaces are mostly longitudinally oriented, $w_{REP,GG}$ is calculated from the area. To compute circumscribed circles, we use the maximum bounding geometry tool implemented in ArcGIS.

In Method 2, we use the maximum inscribed circle to calculate w_{REP} as

$$w_{REP,EE/HH} = \frac{A_{EE}}{d_I} \quad \text{and} \quad w_{REP,GG} = d_I, \tag{2}$$

where $A_{EE/HH}$ is the surface area of the type EE or HH shape and d_I is the diameter of the maximum inscribed circle. Contrary to Method 1, $w_{REP,GG}$ is directly associated with d_I , while EE and HH surfaces are derived from the surfaces respective area. The centre of the largest circle that fits inside an arbitrary polygon is the point that is inside of the polygon and furthest from any point on the edges of the polygon. This largest circle touches the edges of the polygon on at least two points. The algorithm that we used is based on Garcia-Castellanos and Lombardo (2007) and was implemented in QGIS. The algorithm solves this problem iteratively to any arbitrary precision and ensures convergence towards a local maximum of the distance to the polygon edges, as follows:

1. Define a rectilinear search region R from (x_{MIN}, y_{MIN}) to (x_{MAX}, y_{MAX}) ;

2. Create a grid of N_x by N_y nodes in R , where Garcia-Castellanos and Lombardo (2007) use $N_x = N_y = 21$;
3. Of the points inside the polygon, find the point that is furthest from any point on the edge;
4. From that point define a new R with smaller intervals and bounds, where Garcia-Castellanos and Lombardo (2007) reduce R by a factor of $\sqrt{2}$, and repeat from step two to get to any arbitrary precision.

Figure 2 illustrates the application of Methods 1 and 2 to type GG, HH and EE surfaces.

Rather than looking at each sidewalk area on its own, the buffered street axis approach of Method 3 evaluates the total combined sidewalk, walkway and driveway area for each street. The representative width is computed as

$$w_{REP} = 0.5 \cdot \frac{A}{l + 2b} \tag{3}$$

Herein, A is the total sidewalk area (all surface types combined) within the buffer, l is the length of the street axis and b is the buffer size. For the evaluation in this paper, we use a buffer of 20 m around each street axis feature. As shown in Fig. 3, the buffers extend beyond the ends of the street axis to include the intersections with all of their sidewalk areas. This can lead to an overestimation of w_{REP} , particularly for short streets. While 20 m are sufficient to enclose the sidewalks of most streets in Vienna, some broad boulevards—such as the Ringstraße—would require bigger buffers. Because of the absence of data about actual total street width, the analysis therefore has to carefully weigh between missing some sidewalks due to smaller buffer sizes or bigger distortions of w_{REP} for short streets due to bigger buffer sizes.

2.3 Circumference over area

For the large scale sidewalk analysis, we adopt a scale-free approach and plot circumference over area of every surface type (Fig. 4a–c). We call them C–A

Fig. 2 Surfaces excerpt from SIS_F showing sidewalks, walkways and driveways and their circumscribed and inscribed circle diameters and w_{REP}

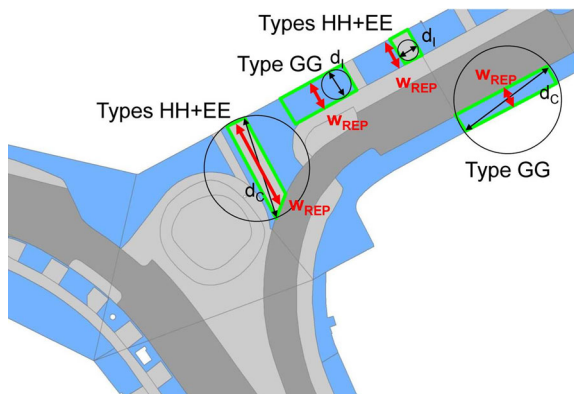
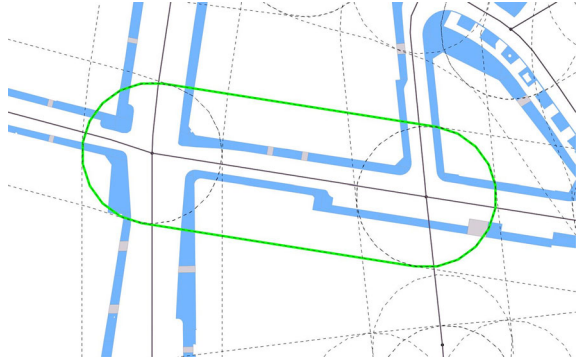


Fig. 3 Buffered street axis with a buffer size of 20 m



diagrams in short. Regular shapes with the shape's circumference U being a function of area A , according to Eq. (4) facilitate the interpretation of the results and are therefore added to the C–A diagrams.

$$U = k \cdot A^\beta. \quad (4)$$

Table 2 provides the parameters for these regular shapes, representing a circle, a square, a rectangle with a side ratio of 1:2 and a rectangle with 1:10. The area's exponent β remains constant at 0.5, the square root obviously.

Figure 4 a–c shows that sidewalk surfaces (GG) are larger than driveway entrances (EE) and walkway entrances (HH) with a tendency to grow narrower with increasing area. This is shown in Fig. 4a by the grey cloud and contour lines moving away from the circle line (red) over the rectangle 2a line (orange) to the rectangle 10a line (purple) with surface sizes around 10^2 m^2 and larger. Figure 4a shows for sidewalk surfaces (GG) that density isolines clearly group along the orange line with the peaks in an area range of $10^{0.5} \text{--} 10^{1.5} \text{ m}^2$, this is from 3 to ca. 32 m^2 .

Figure 4b illustrates that driveway entrances (EE) are bigger in size than walkway entrances (HH, Fig. 4c), as one would expect from door widths. Density peaks in a range around 10^1 m^2 for walkway entrances versus a range between 10^0 and 10^1 m^2 for driveway entrances, which peak around $10^{0.5} \text{ m}^2$ of area and a 10^1 m

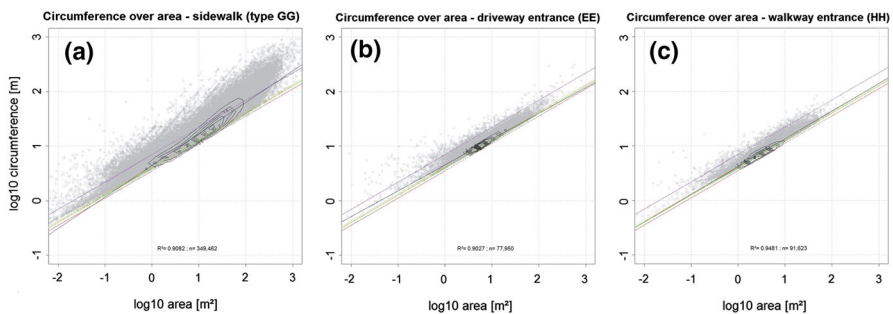


Fig. 4 Circumference-area (C–A) scatter plots in double \log_{10} scale for surface types GG (a), EE (b) and HH (c) with relative density isolines. Coloured straight lines show regular shapes for comparison: circle (red), rectangle 2a (orange) and rectangle 10a (violet) (colour figure online)

Table 2 Parameters of selected regular shapes for equation

Shape	k	β	Colours in Fig. 4
Circle	3.54	0.5	Red
Square	4.00	0.5	Green
Rectangle 2a	4.24	0.5	Orange
Rectangle 10a	6.96	0.5	Violet

circumference. The density spikes clearly group along the rectangle 2a line with no grey circle actually touching the red circle line. Walkway entrance (HH) density groups between the circle (red) and the rectangle 2a line (orange) as well and it can be observed that no grey dot reaches as far as the circle line. Both surface types, EE and HH, are distinctly smaller than GG surfaces and appear to be packed around the rectangle 2a line, whereas higher densities (contour lines) of the GG surfaces reach and cross the rectangle 10a line. In comparison with GG surfaces, EE and HH surfaces show a more compact form, as their density peaks appear between the lines for circles and rectangles with a side proportion of one–two times a . Although, from this point of view, it appears likely to expect that the estimation method for types EE and HH may lead to greater deviations of w_{REP} from the actual sidewalk width, we apply this method nevertheless since the C–A plots indicate that the three applied methods are legitimate and valid. On the one hand side, for GG surfaces, Method 1 is expected to underestimate w_{REP} , because the rectangular-style polygon's area is divided by its diagonal, which is bigger than the longitudinal side. Method 2, on the other hand, is expected to overestimate widths because it relies on the maximum circle. For types EE and HH, the circumscribed method, due to seeking for the polygon's longest diagonal, is expected to overestimate w_{REP} . Method 3 is expected to behave indifferently, as the buffer includes sidewalk surfaces from the intersecting roads and the buffer width is added to street length as denominator in Eq. (3).

3 Results

The mean resulting values and the 15th and 85th percentile of resulting w_{REP} are shown in Table 3. Percentile values were chosen over extremal values (min, max) as the SIS_F database includes surfaces close to zero in area and circumference—see Table 1.

Method 3 aggregates from $n = 519,034$ surfaces to $n = 26,639$ street links. Table 4 summarizes the cumulative results by giving the spans for $w_{\text{REP},15\%}$, $w_{\text{REP},50\%}$ and $w_{\text{REP},85\%}$ values distinguished by administrative districts. The 85th percentile was chosen, as this is a standard percentile in transport measurement and design. For example v_{85} gives the speed that is underrun by 85% of drivers/riders, a parameter often used for road design (FSV 2011, 2014, 2015a, b). The widths exceeded by 85% of measurements are given by $w_{\text{REP},15\%}$. In addition, the array of districts that yield the envelope curve of maximum and minimum w_{REP} cumulative distribution lines is given in the last two columns. Vienna's district numberings runs

Table 3 Distribution parameters of w_{REP} (m) by datasets and methods

Parameter	Method	GG	EE + HH	GG + EE + HH
15th percentile	1—Circumscribed	0.83	2.17	1.05
	2—Inscribed	0.97	1.77	1.19
	3—Buffered street	–	–	1.51
Mean	1—Circumscribed	1.82	3.62	2.41
	2—Inscribed	2.11	3.04	2.42
	3—Buffered street	–	–	4.83
85th percentile	1—Circumscribed	2.67	4.87	3.61
	2—Inscribed	3.06	4.12	3.48
	3—Buffered street	–	–	7.04

Table 4 Ranges of specific parameters of the resulting cumulative distributions

Method	$w_{\text{REP},15\%}$ (m)	$w_{\text{REP},50\%}$ (m)	$w_{\text{REP},85\%}$ (m)	Districts with $w_{\text{REP},\text{MAX}}$	Districts with $w_{\text{REP},\text{MIN}}$
1—Circumscribed	0.51–1.01	1.51–2.26	3.01–4.01	1, 11, 20	14, 15, 23
2—Inscribed	1.00–1.26	1.51–2.76	2.76–4.26	1, 3, 20	13, 14, 23
3—Buffered street	0.26–2.50	1.01–3.25	2.26–4.25	5, 6	13, 22, 23

from the inside out, from 1 to 23:1 is the central business district, 3, 5 and 6 are within the second ring road, while the city's fringe comprises of districts 13 and 14 in the West, 22 in the East and 23 in the South.

As the aggregation of w_{REP} per city district proves to deliver limited further insight, we calculate means of w_{REP} per cell of a 500 by 500 m grid over the city's extents. Figures 5 and 6 illustrate the spatial distribution of mean w_{REP} for Methods 2 and 3, respectively. Method 1 is not depicted since the resulting visual is very similar to Method 2 as shown in Fig. 5.

Visual analysis of the distributions suggests a non-random distribution. We therefore check for spatial autocorrelation using Moran's I analysis (Li et al. 2007). The Moran's I value for method 3 is 0.67, while method 1 and 2 result in a Moran's I of 0.31 and 0.34, respectively. A Moran's I value of -1 indicates perfect dispersion, while $+1$ indicates perfect correlation, with values close to $-1/(n - 1)$ indicating a random spatial pattern.¹ Both z -scores and p values of all three methods indicate spatial autocorrelation that is significant at the 5% level. For Method 3 (buffered street), a highly significant cluster of wide sidewalks emerges in the inner city, while this is not observable for Methods 1 and 2. Figure 7 shows hot and cold spots detected using a local Moran's I analysis of Method 3 results. Dark red cells indicate hotspots where sidewalks are wide in the respective cell as well as in its neighbouring cells. These cluster in the center of the city dark blue cells at the city

¹ <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm>.

Mean wREP

in meters

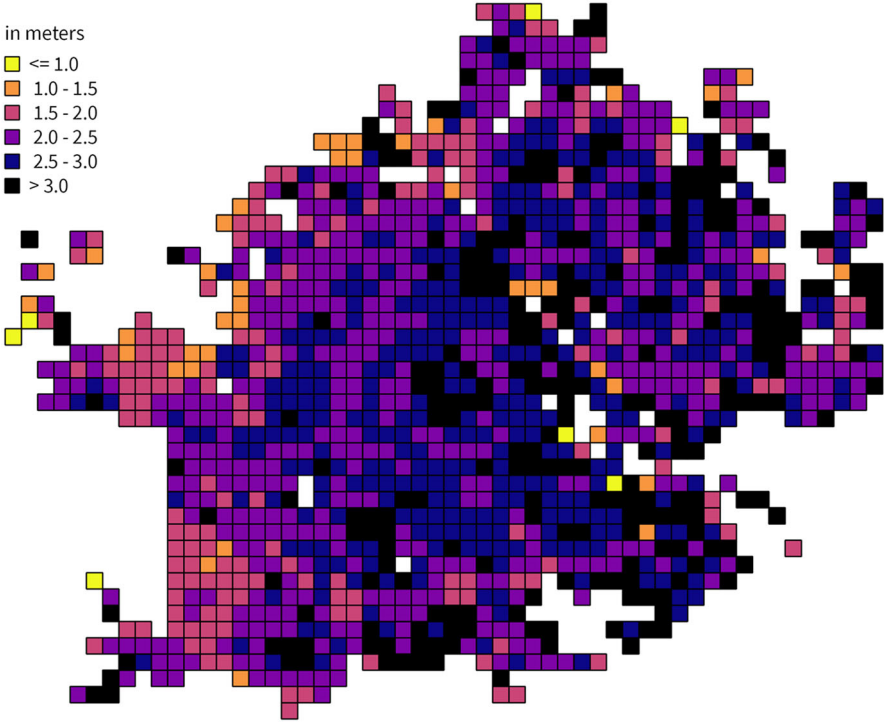
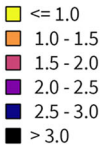


Fig. 5 Mean w_{REP} of Method 2 based on a 500 m grid, types GG + EE + HH

borders, in contrast, indicate areas where cells and their neighbouring cells contain narrow sidewalks.

Type GG surfaces tend to have a different logical orientation as elements of the sidewalk than EE and HH surfaces: GGs are oriented along the sidewalk, while EEs and HHs are usually oriented across. Therefore, we plot cumulative distributions of GG versus EE + HH samples separately for Methods 1 and 2—circumscribed and inscribed (Fig. 8). The difference of Vienna’s total sample (thick lines) at 50% shares is 1.50 m for circumscribed and 0.75 m for inscribed. Table 5 shows the width ranges at specific percentages for comparison between datasets and Methods 1 and 2. Method 2 appears to close in from Method 1, as “black” moves up the scale, while “blue” moves down. Method 3 could not be analysed in such a way by definition, since it does not distinguish between surface types.

4 Discussion

Results are as granular as the input data and method characteristics determine. In our results, we have aggregated all representative sidewalk widths, either by districts or in a grid. These are two options among many, and for other use cases it

Mean w_{REP}

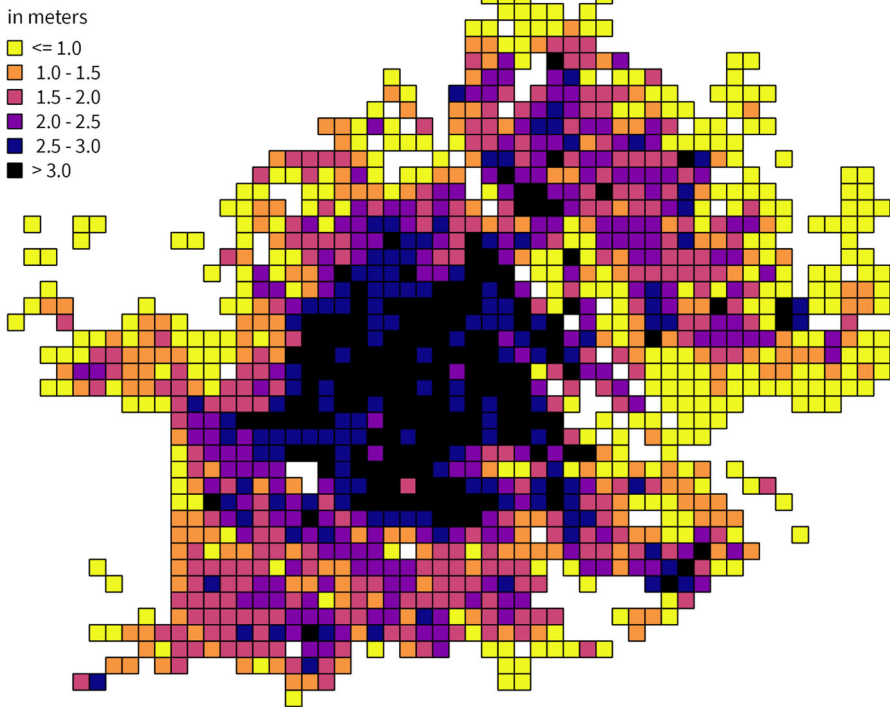


Fig. 6 Mean w_{REP} of Method 3 based on a 500 m grid, types GG + EE + HH

may be more appropriate to aggregate on more detailed levels, such as on a building block or street level or even individual surfaces.

4.1 Context

While the results of our methods agree on some aspects, considerable differences between different method results can be observed as well, which we have presented in graphical and numerical form. Because the cumulative distributions for Vienna (Fig. 8) take a sigmoidal shape, which is often an indicator for a normal distribution, we tested it. Both, the visual evaluation of histograms as well as the Kolmogorov–Smirnov normality test, which is considered suitable for datasets bigger than 2000 specimen, reveals that all three w_{REP} data sets are not normally distributed.

The cumulative distributions of w_{REP} (Fig. 8) reveal a long tail reaching to 10 m of width and beyond. These surfaces do not necessarily represent classic sidewalk settings (as in a cross section of wall–sidewalk–lanes–sidewalk–wall) but represent squares and plazas that are not classified as type FZ (pedestrian precinct). The distributions' long tails are one reason why the normality tests do not reveal a Gaussian distribution of data.

Spatial Autocorrelation - Local Moran's I

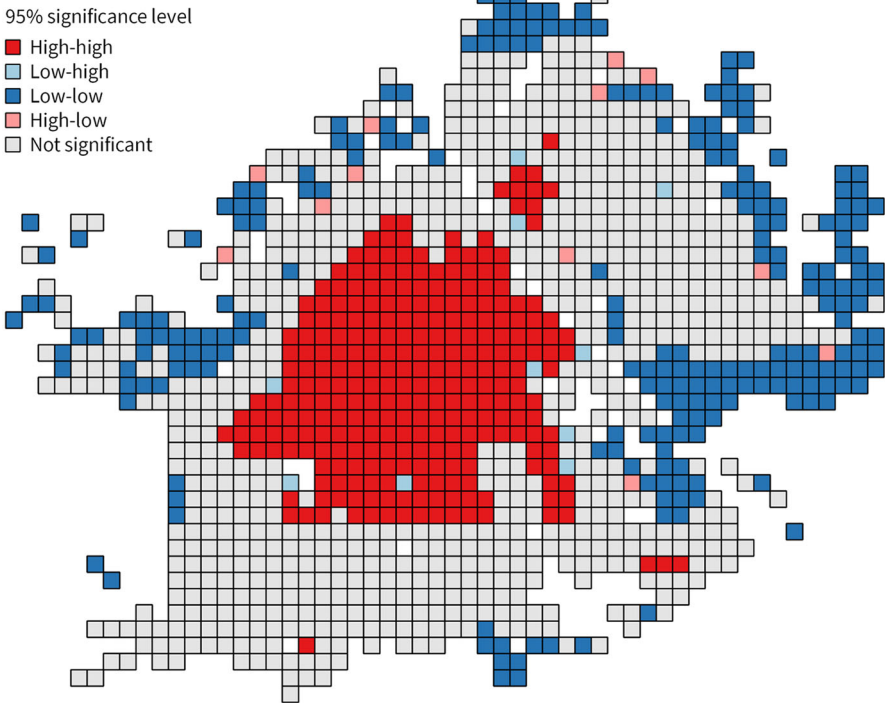


Fig. 7 Local Moran's I spatial autocorrelation of mean w_{REP} of Method 3 based on a 500 m grid. High-high in dark red colours clusters in the city centre, while low-low in dark blue colours is mostly situated at the urban fringe (colour figure online)

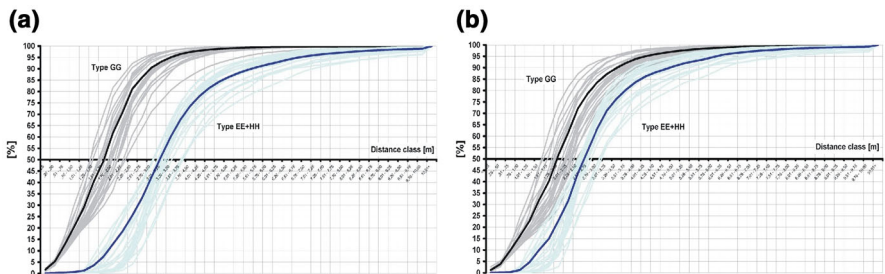


Fig. 8 Cumulative frequency distributions for type GG and types EE + HH of w_{REP} for all 23 districts and Vienna total according to Method 1 **a** with types GG ($n = 349,462$) and EE + HH ($n = 169,572$) and Method 2 **b** with types GG ($n = 351,160$) and EE + HH ($n = 169,572$)

There is a significant difference in results when using Method 3 in comparison with Methods 1 and 2, where the circumscribed method results in the least spreading/diverging curves. While individual district results from Method 3 vary considerably from Methods 1 and 2, spanning a wider range of values, the cumulative curve for Vienna's total datasets does not differ much. District

Table 5 Ranges of specific parameters of the cumulative distributions comparison from Fig. 8

Methods and datasets	$w_{REP,15\%}$ (m)	$w_{REP,50\%}$ (m)	$w_{REP,85\%}$ (m)
Circumscribed GG	0.51–1.00	1.25–2.25	1.76–3.75
Circumscribed EE + HH	1.51–2.75	2.76–3.75	4.00–6.25
Inscribed GG	0.51–1.00	1.26–2.50	2.01–3.50
Inscribed EE + HH	1.01–2.25	2.26–3.00	3.25–5.25

distributions show a relative stability explicated as envelope curves to the set of curves. While district 23, a peripheral district with a lot of industrial areas, consistently appears on the list of minimum sidewalk widths, other peripheral districts (such as 13, 14 and 22) are part of the smallest widths curve sets as well. Maximum widths show more variety. Nevertheless, districts 1 and 20 bound the curve sets from the upper end in two out of three methods, indicating a inter-methodical stability.

Figure 9a clearly shows that, for type GG, the inscribed circle method overestimates w_{REP} in comparison with the circumscribed method. For types EE and HH (Fig. 9b), an opposite relation is observable, and the inscribed method underestimates in comparison with circumscribed. Both observations are in line with expectations, see scheme in Fig. 2. The linear correlation coefficients for EE and HH are smaller than 1 (0.83 and 0.84), while GG surfaces result in a larger than 1 correlation (1.17). All three correlations display very high goodness of fit values (R^2) with 0.844 (HH), 0.896 (EE) and 0.904 (GG). In contrast to GG surfaces, EE and HH surfaces display a rather strict superior boundary at the relation of $w_{REP,I}/w_{REP,C} = 1$ with just a tiny proportion of outliers beyond this virtual border. The relation of circumscribed to inscribed for GG surfaces reveals a larger than 1 coefficient, but does not show any boundary, as the dots appear to scatter more symmetrically around the first median line.

In comparison with single surface consideration in Methods 1 and 2, the buffered street axis method (3) provides an aggregate approach. Surfaces from all three types within the buffer are consolidated and divided by the street buffer’s characteristic length $l + 2b$. For sake of comparison with Methods 1 and 2, we assume most

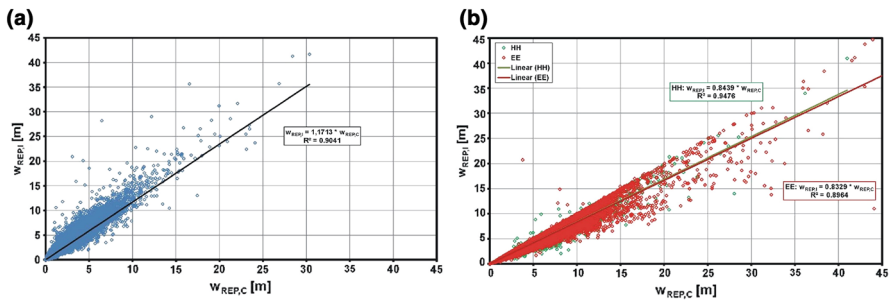


Fig. 9 Scatter plots and linear approximation functions of w_{REP} from Method 2 (vertical axis) versus Method 1 (horizontal axis) for **a** type GG surfaces with $n = 349,462$ and **b** type EE + HH surfaces with $n = 77,950 + 91,623$

streets to have a left and a right sidewalk; therefore, we multiply with 0.5. The use of $l + 2b$ instead of simply l as denominator tends to produce better results since otherwise w_{REP} would be overestimated systematically. This means that circumscribed and inscribed methods favour streets with one wide sidewalk over those with two narrower sidewalks. An improvement of the buffered street axis method would ask for the variation of buffer size within a sensitivity analysis of results. When street link datasets include lane widths and lane numbers, buffer size could be adapted accordingly for every individual link. This would improve the accuracy of results considerably, instead of running a constant buffer size analysis over every type and dimension of roads.

To our knowledge, only two approaches related to ours exist in the literature. One is Freeman and Shapira's (1975) development of the minimum area rectangle (MAR) method for calculating the length and width of a rectangle circumscribing an entity. It is likely that this rectangle's length reveals similar results to our circumscribing circles—Method 1. In the other approach, Zhu and Lee (2008) investigated Moran's I for various transport-related variables. While some variables showed a significant impact, sidewalk width did not show significant clustering.

4.2 Comparison and validation

To further evaluate our results, we compare our methods to an independent reference dataset called graph integration platform (GIP). GIP is the public administration's official street network dataset of Austria.² Each street network link is described in detail, including cross-sectional information with street and sidewalk average and minimum width. The average sidewalk width along a street link l , which is used for the following comparison, is computed as the sum of sidewalk widths w_k over all sidewalks along l —Eq. (5)

$$w_{\text{REP,GIP},l} = \sum_k w_k \cdot (p_{\text{TO}} - p_{\text{FROM}}) \quad (5)$$

where p denotes the percentage to and from and describes the exact location of the modelled sidewalk along the link. Figure 10 presents a scatterplot of w_{REP} according to the buffered street axis method and the width according to GIP. Each data point represents one network link. The observed R^2 is 0.504. The comparison shows that GIP widths tend to exceed results of the buffered street axis method. The biggest errors are due to the following effects: (1) streets where the 20 m buffer only contains parts of the relevant sidewalk areas; (2) differences between GIP and the SIS_F urban surface management database, where GIP tends to be more generous while the SIS_F database tends to classify more areas as non-sidewalks (for example, green areas surrounded by sidewalks).

So far, none of our methods consider the qualitative distinction between a wide one-sided sidewalk and narrow two-sided sidewalks. Even though the resulting

² For the GIP standard documentation, please refer to http://open.gip.gv.at/ogd/gip_standard_beschreibung.pdf.

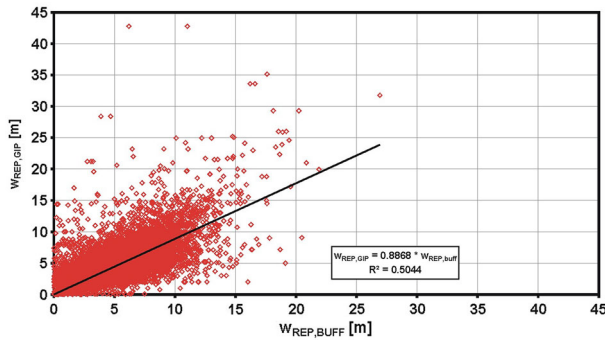


Fig. 10 Scatter plots and linear approximation functions of w_{REP} of buffered street axis and $w_{\text{REP,GIP}}$ ($n = 26,134$)

numbers may be the same, e.g. in Method 3, the qualitative perception by pedestrians may differ considerably.

To get a feeling for the quality of our applied methods, we performed a hands-on plausibility check. For each surface in a selected neighbourhood (three blocks), we manually measured the sidewalk widths in the SIS_F data set. For nearly rectangular surfaces, this was the smallest distance perpendicular to the kerbstones. For irregular surfaces, the distance most plausibly perceived as representative sidewalk width was measured. The plausibility check delivered the following results (see Fig. 11):

- Finding 1: $w_{\text{MEASURED}} - w_{\text{REP}}$ differs most for GG; HH produces the best fit, EE slightly less.
- Finding 2: EE surfaces (driveway entrances) are often wider than the sidewalk width.
- Finding 3: The inscribed method is more exact than the circumscribed method.

Finding 1 is due to the fact that GG surfaced are often irregular, while HH and EE surfaces are mostly perfectly rectangular.

Finding 2 results in an overestimation of w_{REP} as the surface orientation appears contrary to the method's preconditions—in partial contrast to the conclusion drawn in Sect. 2.3. We therefore calculated a corrected width “EE*” with the method used for GG rather than HH which led to a significant increase in fit. Finding 3 is a consequence of the used calculation method where especially in rectangular surfaces the inscribed method renders perfect results for w_{REP} , while the circumscribed method overestimates distances due to the Pythagorean theorem. Given a sufficient number of HH and EE surfaces, this (adapted) method produces a good overview of w_{REP} values.

One further approach for improvement of the presented methodology is to systematically calculate the logic transversal and longitudinal orientation of sidewalk surfaces from neighbourhood relationships between surface types irrespective of actual surface orientation. This is expected to be an advancement benefitting from treating groups of surfaces instead of every surface by itself. Under

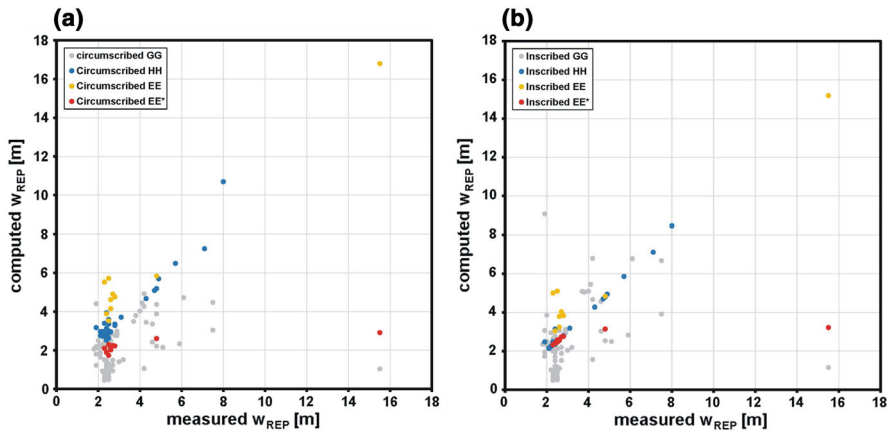


Fig. 11 Scatter plots of measured w_{REP} with computed w_{REP} according to Method 1 **a** and Method 2 **b** for types GG ($n = 116$), HH ($n = 36$), EE ($n = 10$) and EE* ($n = 10$)

ideal circumstances, a GG type surface is expected to have its longer extension oriented with the longitude of the sidewalk, while for surfaces denoting the access to buildings over the sidewalk, such a clear distinction between width and length in terms of pedestrian's logic is not possible from geometry alone.

Sidewalk surfaces, especially in intersection setups, are segmented by lines that result from virtual alignment lines between house corners. These fragmentations reduce surface size artificially and are likely to lead—especially for Method 2—to an underestimation of w_{REP} . For example, when a square is divided in two isosceles triangles, the relation of areas is 2, but the relation of inscribed diameters changes to 1.707. Circumscribed diameters remain the same.

4.3 Requirements

Measurements of the actual space that pedestrians demand for (Schopf 1985), put our results into the following perspective: two-person encounters ask for a width of 2.21 m at 50% share and 2.56 m at 85% share. While with the circumscribed method the 50% value is met only by the better half of districts, the 85% value is easily overachieved. A similar situation is present with the inscribed method. The buffered street axis method results in show that the total dataset meets the 50% as well as the 85% requirements. The Austrian transport engineering guidelines on sidewalk design (FSV 2015a, b) give 1.50 m as the minimum width for enabling the meeting of two opposite walking pedestrians and ask for a basic sidewalk width of 2.0 m. All three methods result in a $w_{REP,50\%}$ that meets this requirement, indicating that only selected districts assessed using the inscribed method ($w_{REP,15\%}$ up to 1.51 m), do not meet the quality standard for sidewalk infrastructure (Table 3).

5 Conclusion

In this paper, we have used three different methods for calculating a representative sidewalk width from a detailed city-wide urban surfaces GIS database.

Our paper contributes insofar as it introduces three methods which can be used to quickly estimate representative sidewalk widths from existing large scale urban surface GIS data. Moreover, the methods' quality, significance and representativeness are tested in a series of critical examinations. In one case, surface types EE, the comparison with manual measurements suggest to change the calculation algorithm accordingly when applying Methods 1 and 2 as these surfaces appear to be more oriented like GG surfaces than HH surfaces.

Results are a quantitative basis for subsequent in-depth analysis. Not only can these results be precisely compared to guideline requirements, they can also evaluate in terms of spatial autocorrelation. The question "Do clusters of extra wide or extra narrow sidewalks exist?" can be answered for areal entities. We performed this examination based on a 500 by 500 m grid to show that Method 3 leads to such a clustering of likewise results.

The task, which cannot be satisfied with our methods, but is one of very high importance to real pedestrian quality, is the consideration of net sidewalk widths. Gross width encompasses the sidewalk surface from wall to kerb, while net width is determined by obstacles and hindrances such as traffic sign posts, bargain baskets, light poles and advertisements cluttering sidewalks. As Fig. 2 shows in its left lower corner, the database already includes some of these fixed obstacles: tree bases and traffic light control boxes. Other fixed obstructions are not included in the present data base. To our knowledge, only lighting post GIS data exists in Vienna to a level precise enough to be included in our kind of calculations—a task open for future endeavours. One pure geometrical approach would be to cookie-cut these obstructions from its underlying sidewalk surfaces. Yet, most of these obstructions such as lamp poles and traffic signs engage only a very small horizontal space, hardly impacting on surface size. A lamp pole's obstruction effect lies in its local space constriction—a singularity in comparison with sidewalk surfaces. Although the road engineering rules pose a car-lane envelope driven demand for putting traffic signs up to 60 cm into the sidewalk surface, systematic surveys of how much these obstacles reduce gross widths have not been undertaken so far to our knowledge. But, we consider research on this issue as dearly needed. In a first approach, undergrad students at our Institute of Transportation are now conducting preliminary enquiries addressing net and gross sidewalk width discrepancy.

In general, GG surfaces due to their variety in shapes and sizes promise lower accuracy than EE and HH surfaces. But how representative are EE and HH surfaces—in their appearance discrete information—in contrast to the larger GG surfaces, representing most of the sidewalk surfaces? Therefore, the question of weighting of method's results in aggregate analysis needs to be targeted for method improvement. As of now, every derived w_{REP} value weighs in equally. However, a weighting method for calculating w_{REP} for road sections or building blocks has to be defined. Shall GG based results are forfeited for the more precise but likely less

representative EE- and HH-based values? It appears conceivable to the authors that when using GG, EE and HH-based results, the individual surfaces size ought to be included as a weight in calculating aggregated representative values, e.g. for a block, a street section or an administrative unit. The question, when is it sufficient to just arithmetically average EE and HH values to achieve an acceptable degree of representation, poses a research task in need of being addressed. We see adding improved methods for dealing with all walkable surfaces—such as pedestrian precincts and separated walkways—as one additional horizon of improvement.

For the methods presented, we envision the following application examples:

1. Automatic enrichment of street network databases with sidewalk width information (for example for the official Austrian reference network GIP as an alternative to extensive manual data collection).
2. City-wide systemic monitoring of sidewalk widths—as absolute values or as a proportion of the total streetscape.
3. Input data for surveys, e.g. Frank et al. (2007).

Thus, we consider our three methods as quick tools for enabling a better understanding of urban systems at the crossroads of transport and surface allocation—thus a step towards leaving the yardstick behind.

Acknowledgements Open access funding provided by TU Wien (TUW). The SIS_F database is courtesy of Vienna city administration's department for roads. A part of this study was funded by Wiener Linien (TB and UL). No influence was taken by the owners and funders on choice of methods and conclusions derived. Thanks to Takeru Shibayama and Bernadette-Julia Felsch for proof-reading and suggestions. Our thanks as well belong to two anonymous reviewers who helped shaping the final version of the paper.

Compliance with ethical standards

Conflict of interest All authors declare no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agampatian R (2014) Using GIS to measure walkability: a case study in New York City. Royal Institute of Technology (KTH), Stockholm
- Dewulf B, Neutens T et al (2012) Correspondence between objective and perceived walking times to urban destinations: influence of physical activity, neighbourhood walkability, and socio-demographics. *Int J Health Geogr* 11(1):1–10. doi:10.1186/1476-072X-11-43
- Duncan DT, Castro MC et al (2012) Racial differences in the built environment-body mass index relationship? A geospatial analysis of adolescents in urban neighborhoods. *Int J Health Geogr* 11:11. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3488969/>
- Duncan DT, Piras G et al (2013) The built environment and depressive symptoms among urban youth: a spatial regression study. *Spat Spatio-temporal Epidemiol* 5:11–25. <http://www.sciencedirect.com/science/article/pii/S1877584513000105>

- Ewing R, Schroeder W et al (2004) School location and student travel analysis of factors affecting mode choice. *Transp Res Rec* 1895(1):55–63
- Frackelton A, Grossman A et al (2013) Measuring walkability: development of an automated sidewalk quality assessment tool. *Suburb Sustain* 1(1):4. doi:10.5038/2164-0866.1.1.4
- Frank LD, Schmid TL et al (2005) Linking objectively measured physical activity with objectively measured urban form. *Am J Prev Med* 28(2):117–125. <http://www.ajpmonline.org/article/S0749-3797%2810%2900297-7/fulltext>
- Frank LD, Saelens BE et al (2007) Stepping towards causation: do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Soc Sci Med* 65(9):1898–1914. <http://www.sciencedirect.com/science/article/pii/S0277953607003139>
- Freeman H, Shapira R (1975) Determining the minimum-area encasing rectangle for an arbitrary closed curve. *Commun ACM* 18(7):409–413. <http://cacm.acm.org/magazines/1975/7/11548-determining-the-minimum-area-encasing-rectangle-for-an-arbitrary-closed-curve/abstract>
- FSV (2011) RVS 03.02.13: Nicht motorisierter Verkehr—Radverkehr. Wien, FSV—Österreichische Forschungsgesellschaft Straße, Schiene, Verkehr
- FSV (2014) RVS 03.03.23: Linienführung und Trassierung. Wien, FSV—Österreichische Forschungsgesellschaft Straße, Schiene, Verkehr
- FSV (2015a) RVS 02.02.37: Geschwindigkeitsbeschränkungen. Wien, FSV—Österreichische Forschungsgesellschaft Straße, Schiene, Verkehr
- FSV (2015b) RVS 03.02.12: Nicht motorisierter Verkehr—Fußgängerverkehr. Wien, FSV—Österreichische Forschungsgesellschaft Straße, Schiene, Verkehr
- Garcia-Castellanos D, Lombardo U (2007) Poles of inaccessibility: a calculation algorithm for the remotest places on earth. *Scott Geogr J* 123(3):227–233. <http://www.tandfonline.com/doi/full/10.1080/14702540801897809>
- Iseki H, Tingstrom M (2014) A new approach for bikeshed analysis with consideration of topography, street connectivity, and energy consumption. *Comput Environ Urban Syst* 48:166–177. <http://www.sciencedirect.com/science/article/pii/S0198971514000891>
- Li H, Calder CA et al (2007) Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geogr Anal* 39(4):357–375. doi:10.1111/j.1538-4632.2007.00708.x
- Lin J-J, Chang H-T (2010) Built environment effects on children's school travel in Taipei: independence and travel mode. *Urban Stud* 47(4):867–889
- Mesbah M, Currie G et al (2012) Spatial and temporal visualization of transit operations performance data at a network level. *J Transp Geogr* 25:15–26
- O'Sullivan D, Morrison A et al (2000) Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach. *Int J Geogr Inf Sci* 14(1):85–104
- Sallis JF, Saelens BE et al (2009) Neighborhood built environment and income: examining multiple health outcomes. *Soc Sci Med* 68(7):1285–1293. <http://www.sciencedirect.com/science/article/pii/S0277953609000318>
- Schneider RJ, Patten RS et al (2005) Case study analysis of pedestrian and bicycle data collection in US communities. *Transp Res Rec* 1939:77–90
- Schopf JM (1985) Bewegungsabläufe, Dimensionierung und Qualitätsstandards für Fußgänger, Radfahrer und Kraftfahrzeuge. Fakultät für Bauingenieurwesen. Wien, TU Wien
- Wong BY-M, Faulkner G et al (2011) GIS measured environmental correlates of active school transport: a systematic review of 14 studies. *Int J Behav Nutr Phys Act* 8:1–22. doi:10.1186/1479-5868-8-39
- Zhu X, Lee C (2008) Walkability and safety around elementary schools. *Am J Prev Med* 34(4):282–290. doi:10.1016/j.amepre.2008.01.024