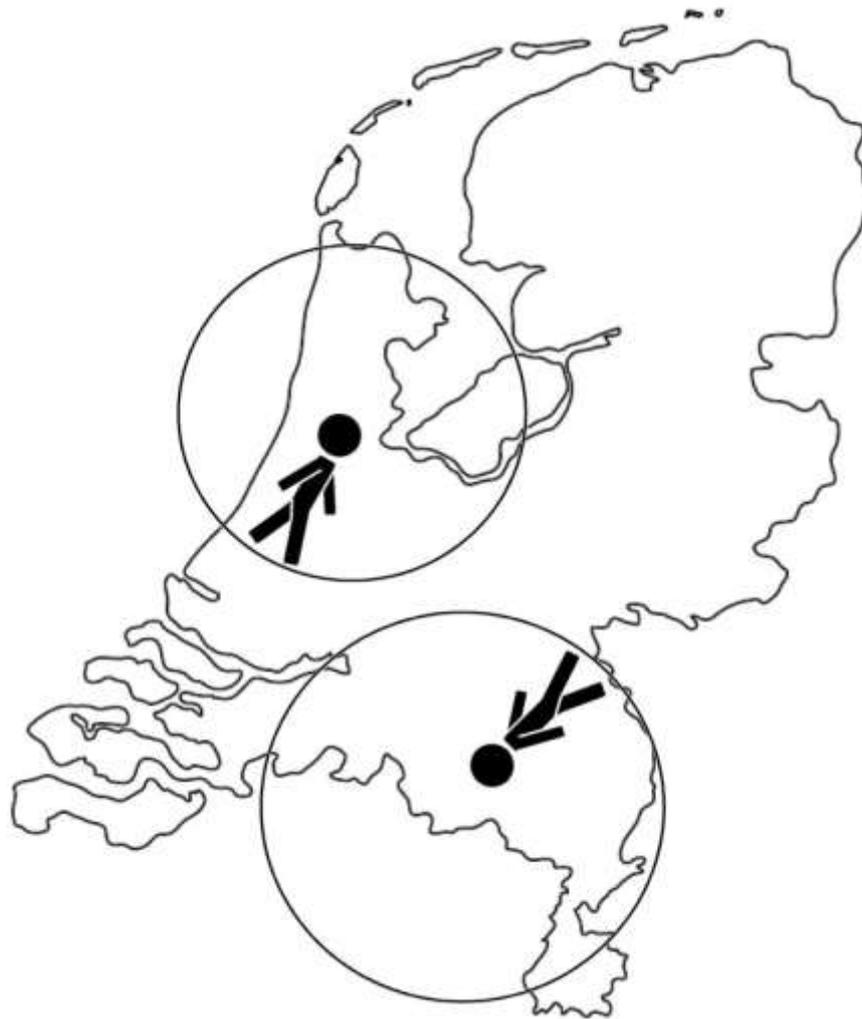


Walkability score Netherlands version 1.0

Documentation of data and methods



Alfred Wagtendonk, Jeroen Lakerveld (August, 2019)
Department of Epidemiology and Biostatistics
Amsterdam Public Health Research Institute
Amsterdam UMC



Walkability score Netherlands version 1.0; Dataset and technical documentation of data and methods

Authors: Alfred J. Wagtendonk, Jeroen Lakerveld

Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam UMC

Project: Geoscience and Health Cohort Consortium (GECCO)

Report date: 6 September 2019, update 25 March 2022

1. Contents

- 2. Introduction..... 4
- 3. Population density..... 7
- 4. Density of retail and service destinations 10
- 5. Land-use mix..... 16
- 6. Street connectivity (intersection density) 20
- 7. Green space density 24
- 8. Side walk density 25
- 9. Walkability index 27
- 10. Discussion 28
- 11. References..... 30

2. Introduction

In this document we document data sources and explain the decision criteria and geoprocessing steps undertaken to construct the first version of a walkability score for different walk ranges around residential address locations in the Netherlands. Starting point for the development of this index was the document 'Walkability v1.2.docx', that was discussed in February/March 2019 within the 'Walkability working group', a collaboration between the 'GECCO' project (VU University Medical Center, Amsterdam UMC) and the Global and Geo Health Data Centre (GGHDC), Utrecht University.

A walkability score can serve different purposes. As emphasised by Shashank and Schuurman (2019) walkability indices and scores have been developed from different disciplines with often different types of walkers and research purposes in mind. The selection and weighting of variables in walkability indices depends heavily on how walkability is defined and needs therefore to be considered carefully. From an overview by Shashank and Schuurman (2019) of definitions of walkability in the most widely cited articles on walkability for health research, we choose the definition by Leslie et al. (2007, p. 113): "The extent to which characteristics of the built environment and land use may or may not be conducive to residents in the area walking for either leisure, exercise or recreation, to access services, or to travel to work". Furthermore, also the exact purpose and function of each variable included in a walkability index needs clarification.

In this project the principle objective of the developed walkability score is to relate it to health behaviours and health outcomes in a potentially wide range of epidemiological research. Central in such research is the precise relation between the conduciveness of the built environment and the physical walking activities of its inhabitants. In our case this can be walking for any purpose, both utilitarian and recreational walking. Important to realize however is that walking scores are associated to recorded physical activity and health outcomes of specific respondents in participating cohorts. From these respondents only the residential address coordinates or postcode or neighbourhood are known. This means that associated walking activities are always linked to a residential location and not to, for example, walking activities on a work or recreational location further away from home. As a consequence the walkability score would need to address the full spectrum of incentives for walking from and/or to a residential home address for potential pedestrians from all possible age classes, health classes and cultural backgrounds.

Obviously, there is no single walkability index that can address multiple epidemiological research goals and pedestrian types at the same time and therefore our main product is in reality not a fixed score but a set of key factors / components for different walk ranges that can be combined in different ways with different weighting schemes to produce a walkability index / score that meets a specific research objective.

Walkability is a composite measure (index) containing a number of environmental characteristics. As was mentioned in the Walkability document by the Walkability group there is no set amount or type of characteristics that 'need' to be included, although the following characteristics frequently appear in walkability indices: Population or residential density, proximity or density of retail and service destinations, land-use mix and street connectivity.

To understand why these factors are relevant and to what extent an environment is conducive or friendly to walking it makes sense to go back in time before the massive introduction of motorized transport in which people still had to rely on walking or on public transportation to get from place to place (Hajna et al., 2015). These circumstances resulted in built environments that were 'naturally' conducive for walking. In general people used to live close to each other and in close proximity to their daily necessary services. Furthermore, a highly connected street network allowed for easy access to these services. Also current residential areas suitable for utilitarian walking (walking for specific purposes such as travel to work or school) are expected to resemble these natural grown historical walking areas, thus areas that are defined by relative high densities of population and service destinations, mixed land uses and high street connectivity.

In the current time we can expect that the car dominated transport system has a large (negative) impact on walking conductivity if there are no safe ways alternatives to walk the streets. We think therefore that the presence and density of a sidewalk infrastructure is another influencing factor for walkability. Finally, for recreational walking purposes we expect that the presence and density of green space (such as parks and forests) and other attractive landscape elements has a positive influence.

We agreed therefore to base the index on the following 6 components:

1. Population density
2. Density of retail and service destinations (retail environment)
3. Land-use mix
4. Street connectivity (intersection density)
5. Green space
6. Side walk presence

All these factors are part of walkability studies and we will come back to them separately for each factor.

Other additional factors from the walkability literature mentioned in the definition document are:

- Safety from crime
- Traffic safety
- Traffic volume and speed
- Pedestrian crossing availability
- Aesthetics
- Air quality
- Shade or sun in appropriate seasons
- Street furniture
- Wind conditions
- Specific walking destinations such as light rail stops and bus stops (Brown et al., 2009)
- Job density

Although relevant, we have decided to not to incorporate these components in the current version of the walkability index. However, one or more of these factors can be considered for inclusion in future versions of the walkability index.

The walkability score or index for each address location in the Netherlands is calculated as a continuous score between 0 and 100 (from low to high walkability) for multiple exposure areas / walk ranges and years dating back to 1989. The score is based on six different spatial components that we selected from other walkability studies and we consider most influential for the spatial variation of walkability in the Netherlands. The different spatial components are computed as raster cell values in a regular spaced grid covering the Netherlands with raster cells of 25x 25 meter. The individual raster cell values for each component are computed using focal statistics (spatial statistics / summaries) for different years and 9 different sized buffers around each raster cell. In this type of GIS computation a so called ‘moving window analysis’ is applied in which for each raster cell the same calculation (e.g. sum of mean of raster cell values in the window) is carried out that is based on the values of the surrounding raster cells in a buffer of a specific form and size (most often a circular buffer). Figure 1 shows an example of a processing cell with its neighbourhood (left figure) and the concept of the moving analysis window (right figure) where the analysis window moves to the right from processing cell A to B.

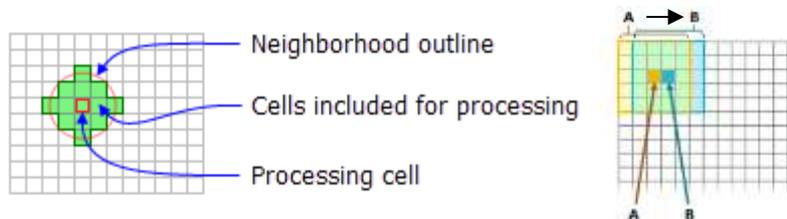


Figure 1 Concept of moving window analysis in focal statistic / neighborhood analysis in GIS

The buffers or ‘neighborhoods’ⁱ concern the size and shape of the area around each raster cell that we consider for walking (the walk ranges), respectively a circular area with a radius of 150-, 250-, 350-, 500-, 750-, 1000-, 1650 and 2000 meter (approximately walking distances between 2 and 30 minutes).

The component values calculated this way for each 25x25 meter location in the Netherlands are extracted and linked to all, approximately 9 million address locations in the Netherlands together with the XY coordinates of the accompanying address location. Because this extraction process takes too much time in a conventional ArcGIS-Python script (about an hour for each of the more than 500 produced variables) we have commissioned the company ‘ObjectVision’ to develop a dedicated application with their GeoDMS solution, which is able to produce these 500+ variables in a few hours only.

As well we extract and link these component values to the centroids of all 452.841 PC6 areas (year 2018) in the Netherlands. Note that PC6 areas differ in size and number of addresses per PC6 area. Especially in the rural areas, PC6 areas can be large and number of addresses small. This means that in general the variable value at the centroid location is less representative compared to the centroid

ⁱ Notice that the use of the word ‘neighborhood’ as used in the ESRI / ArcMap software documentation in its neighborhood analysis functions can be confusing, as these neighborhoods refer to uniform zones (e.g. circular or square) around the processing cell while the term ‘neighborhood’ is in common speak also used to indicate a certain spatial level of not uniform administrative zones. In this document the term ‘neighborhood’ can refer to both meanings, but where this is not immediately clear from the context we try to make this more clear, e.g. by talking about administrative or circular neighborhoods. However, with the term ‘neighborhood analysis’ we always refer to the uniform neighborhoods used in the spatial analysis functions in ArcMap from ESRI.

value in small PC6 areas in more urban areas. Alternatively, we can therefore provide statistical summaries of the component values for specific or all different buffer radii for all individual PC6 zones using the zonal statistics function in ArcMap. Summarizing values over PC6 zones instead of extracting variable values at PC6 centroid locations makes especially sense for variables of higher spatial accuracy and to get better estimates in rural areas.

Furthermore, statistical summaries are calculated or reported per administrative neighbourhood (1988, 1993, 1995, 1997, 1999, 2001, 2003 – 2018) and PC4 area (1998 – 2017). Also here we can provide statistical summaries of the component values for specific or all different buffer radii summarized to neighborhoods and/or PC4 zones of one or multiple years using the zonal statistics function in ArcMap. This approach has the advantage that the summarized buffer values in the neighborhoods include information from adjacent neighborhoods / postcode zones and overcome one of the issues in the Modifiable Unit Area Problem (MAUP)ⁱⁱ.

The spatial components that are input for the index on each location were created in a Geographical Information System (GIS software: ArcGIS desktop 10.6 from ESRI) involving spatial data processing and analysis of raster and vector map layers. Because these GIS operations have to be repeated for a large number of years and walk ranges we automated them using Python scripts that apply the ArcPy modules of ArcGIS.

The walkability index is subsequently calculated for each address location in SPSS by summing the normalized z-scores of each of the six components or measures after which the result is normalized to a value between 1 and 0. The different steps and the computation of each component are explained in detail in the following sections 2 to 7.

3. Population density

Introduction

Population density is considered an important component of the walkability metric as densely populated areas make walking efficient, decrease the appeal of driving through congested areas where parking is often scarce, and creates demand for destinations (Brown et al., 2009).

Data selection and preparation

As mentioned in the walkability definition document, population density can be described by the number of inhabitants per km². These data are registered by and available from Statistics Netherlands (CBS StatLine) for every year until about 50 years back.

More precisely, CBS offers these data on different administrative scale levels from the national level to the neighbourhood level and for two specific years also on the PC6 levelⁱⁱⁱ. Moreover, to increase

ⁱⁱ MAUP is a well-known spatial problem first described by Openshaw in 1984, see e.g. Buzelli (2020) for an overview and possible ways to deal with the problem.

ⁱⁱⁱ For the years 2012 and 2014 CBS offers a complete database on PC6 level with population data, which concerns however 'maatwerk' with a price-tag, see further: <https://www.cbs.nl/nl-nl/maatwerk/2015/46/kerncijfers-postcodegebieden-2014-2012>

the comparability of population data through the years, CBS offers also a gridded product in which population data is allocated to fixed grid cells of respectively 100 and 500 meter resolution: the so called 'vierkantstatistieken', see the map example in Figure 2. Main advantage of this dataset is its relatively high resolution and that the position and dimensions of the grids remains stable over time (in contrary to many administrative areas and borders).

To create the population density component at address and PC6 level we use the CBS dataset '100 meter Vierkantstatistieken' for the years 2000 to 2017, where the attribute field 'INWONER' contains the number of inhabitants per 100x100m gridcell / hectare.



Figure 2 Map example CBS vierkantstatistieken 100 meter

The component is created for 18 different years, 8 different neighborhood radii and 3 different administrative neighborhoods. For the years prior to 2000 there are no '100 meter Vierkantstatistieken' available so we use population data from the CBS 'wijk- en buurtkaarten' instead, which is available for the years 1995, 1997 and 1999. In these data population density values are given for each neighborhood. Consequently, the spatial accuracy is much lower for these years compared to the data from 2000 onwards.

To create the population density component at PC4 and neighborhood level we use the CBS statistical data series for PC4 areas and the 'Wijk- en buurtkaarten' in which an attribute with population density is already provided.

The CBS data is available from the sources listed here below.

CBS data Vierkantstatistieken :

<http://geodata.nationaalgeoregister.nl/cbsvierkanten100mv2/extract/cbsvierkanten100m20002014v1.zip>

<http://geodata.nationaalgeoregister.nl/cbsvierkanten100mv2/extract/cbsvierkanten100m2015v1.zip>

<http://geodata.nationaalgeoregister.nl/cbsvierkanten100mv2/extract/cbsvierkanten100m2016v1.zip>

<http://geodata.nationaalgeoregister.nl/cbsvierkanten100mv2/extract/cbsvierkanten100m2017v1.zip>

CBS data 'Wijk en buurtkaarten' 1995 to 2017 via WFS service link within Q-GIS:

<https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/effe1ab0-073d-437c-af13-df5c5e07d6cd> (buurten)

Neighborhood statistics (georeferenced tables without maps) can also be downloaded separately via <https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/wijk-en-buurtstatistieken>

The postcode 4 units with population data from 2015 to 2017 by CBS are downloaded from:

http://geodata.nationaalgeoregister.nl/cbspostcode4/extract/CBS_PC4_2015.zip

http://geodata.nationaalgeoregister.nl/cbspostcode4/extract/CBS_PC4_2016.zip

http://geodata.nationaalgeoregister.nl/cbspostcode4/extract/CBS_PC4_2017.zip

(see also <https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische%20data/gegevens-per-postcode>)

The postcode 4 population data from 1998 to 2014 by CBS are downloaded from:

CBS Statline: <https://statline.cbs.nl>

To get the right data type '*Bevolking en huishoudens; viercijferige postcode, 1 januari 1999*' in the search box and change the year accordingly to the year you are looking for. Next choose: '*Pas selectie aan*' and select only the categories you need. Next, the table data of the different years is linked to the CBS PC4 unit map of 2015 as CBS does not provide PC4 unit maps of earlier years.

GIS operations for production of component maps and table variables

A summary of GIS operations carried out to produce the required table variables is presented here below. For exact procedures we refer to the Python scripts used to automate these GIS operations.

Adress/PC6 level

Python script: 1_Population_density_Walkability_vierkantstatistieken.py

- Convert 100x100 meter CBS buurtstatistieken (2000-2017) and CBS buurtkaarten polygons (1995, 1997, 1999) to 25x25 meter raster cells (value: popdens / ha)
- Calculate neighbourhood statistics (focal statistics/ mean) for different neighbourhood radii and different years. Resulting values are population densities per neighbourhood radius in hectare (ha) per year.
- Calculate z-score for each raster cell (raster cell value - Mean) / SD)

Python script: 1_Population_density_extract_values_2_PC6.py

- Extract values to PC6 centroid table for each exposure variable

Geo-dms application with adapted lines in 'stam.dms'

- Extract values to address coordinate table for each exposure variable

PC4/Neighborhood level

Python script: 1_Population_density_Walkability_administrative_units.py

- Produce results on neighborhood and PC4 level by exporting available population densities in CBS neighborhood and postcode 4 maps from 1995 and onwards to the desired table format

Notes:

1. The input tables contain gridcells with NoData values expressed as negative numbers (-99997). Where NoData occurs this means the attribute information for one or more fields of a gridcell is classified secret / private. For the field 'Inwoners' this is the case if the cell represents an area with less than 5 inhabitants. In our procedure these negative values are replaced by a value of zero to prevent wrong population densities per neighborhood radius.

2. Some areas are missing because they exist as address locations but don't occur in the CBS vierkant-statistieken, this is e.g. the case for all recreation dwellings in the Netherlands (e.g. take a look at the 'vakantieparken' on Ameland)

In Figure 3 a map example is given with Z-scores for the year 2015 within a 1000 meter walk range.

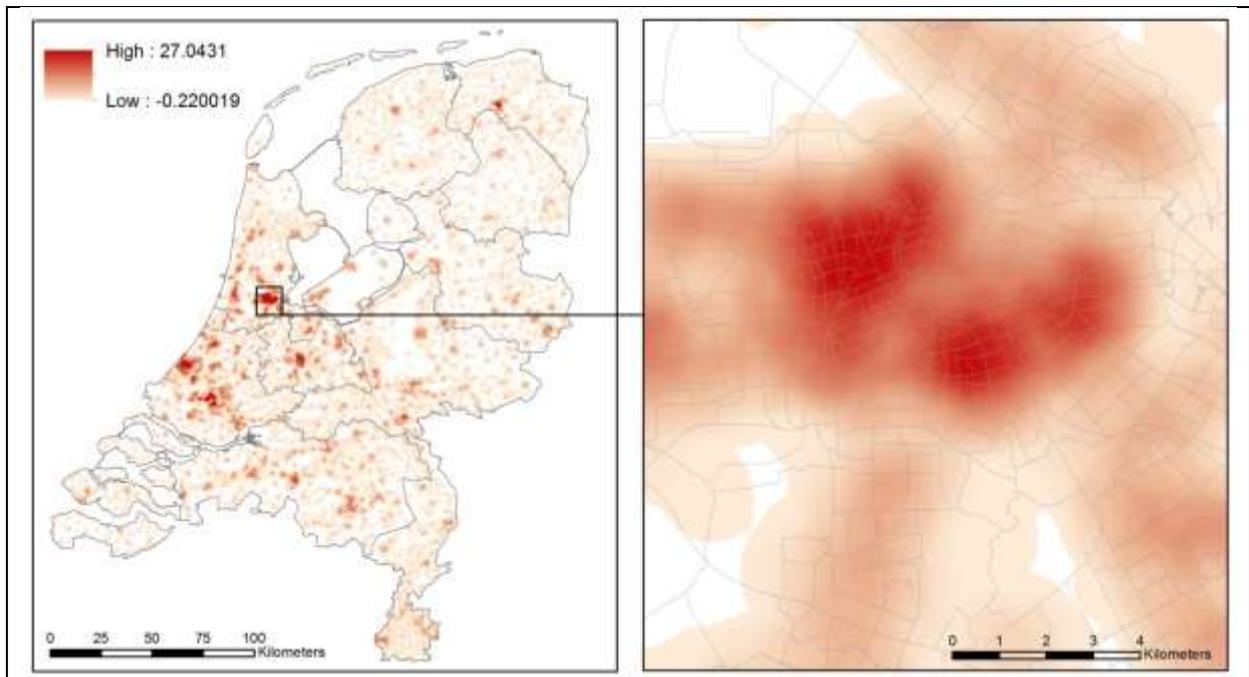


Figure 3 Map example: Z-scores population density 2015, 1000 meter radius

Discussion

We consider population density an important factor for any walkability study. We have used yearly CBS population data on the most detailed spatial level (100 x 100 meters resolution) for the years 2000 to 2017. We consider these CBS data a fairly good data source for our purposes, but unfortunately data before 2000 is only available on a neighbourhood level for the years 1995, 1997 and 1999.

For specific studies that involve only certain years it is also possible to use CBS population data on a PC6 level. This concerns the years 2012, 2014, 2015, 2016 and 2017. This is however data with a price-tag.

Alternatively it is also possible to use residential density instead of population density. However, we choose for population density because for residential density we do not have access to historical data before 2015. For studies concerning the period 2015 or later it can be used however, appendix 2 describes a possible procedure.

4. Density of retail and service destinations

Introduction

Retail and other service destinations are often considered the glue that hold walkable urban places together. They attract people to the street and, subsequently, people together provide more safety,

more vibrancy and attract again more people. Net retail area ratio is a typical component of walking indices such as the Walkability Index ([Frank et al., 2010](#)) and the presence and proximity of retail and other service destinations has e.g. been positively associated with utilitarian walking in 80% of studies reviewed by [Sugiyama et al. \(2012\)](#).

Data selection and preparation

Although (commercial) datasets are available with a higher spatial and attribute resolution we have chosen to use an alternative with less spatial accuracy, but with better temporal accuracy: the 'BBG bodemgebruiksbestanden' available for the years 1989, 1993, 1996, 2000, 2003, 2006, 2008, 2010, 2012 and 2015^{iv}. All available files are in vector format.

The files for 1989 and 1993 have been acquired by specific request to CBS (Hans Visser: j.visser@cbs.nl). Source data for the years 1996, 2000, 2003 and 2006 was acquired from DANS-KNAW (Data Archiving and Networked Services). Source data for the years 2008 to 2015 was downloaded from:

<http://geodata.nationaalgeoregister.nl/bestandbodemgebruik2008/extract/bestandbodemgebruik2008.zip>

<http://geodata.nationaalgeoregister.nl/bestandbodemgebruik2010/extract/bestandbodemgebruik2010.zip>

<http://geodata.nationaalgeoregister.nl/bestandbodemgebruik2012/extract/bestandbodemgebruik2012.zip>

<http://geodata.nationaalgeoregister.nl/bestandbodemgebruik2015/extract/bestandbodemgebruik2015.zip>

Instead of only calculating a retail ratio based on the BBG class 'Detailhandel en horeca', we decided to add social-cultural service destinations as well ('Sociaal-culturele voorzieningen'). This class includes destinations like schools, universities, conference centres, day-care centres, cultural and neighborhood centres, hospitals, nursing homes, general practitioners, sanatoria, medical services, religious buildings, museums and concert halls.

An important drawback of using the BBG land use maps which we discovered during the process, is that from 2008 onwards new data from Locatus on retail has been used to improve the land use maps, which was not available for older years (personal information Hans Visser / CBS, see e-mail d.d. 5 juni 2019). We noted for example large differences in the class 'detailhandel' between 2008 and older maps for familiar shopping areas in Amsterdam. In practice this means that many relatively small retail areas that are present on the land use maps from 2008 onwards are not mapped on earlier versions of the land use maps, while most of these areas were of course already present. To correct for these differences we made the following assumptions:

- established retail areas do not change often (and are part of municipal zoning plans)
- the majority of new retail areas are established in urban expansion areas

On the basis of these assumptions we started with the retail area on the BBG map of 2008 and used the CBS maps 'bevolkingskernen' (showing morphological continuous areas that are mainly inhabited

^{iv} CBS statistics also produced land use statistics for the years 1989 and 1993. However, these files should be requested at CBS separately and are not harmonized with more recent years.

by people) available for the years 2001, 2008 and 2011 to determine the urban expansion areas between 2001 and 2008. Subsequently, we used these areas to remove the more recent retail area in the BBG land use maps of 2000 and 2003. Before 2001 there are no CBS maps 'bevolkingskernen' available, but the Dutch environmental assessment agency (PBL) has produced generalized urbanization maps for several years (among others 1980, 1989 and 1993) which we used in a similar manner to correct the land use maps of 1989 and 1993. The following overview shows the corrections made for each land use map.

- BBG land use 1989: retail class corrected with urban expansion 1989 – 2008
- BBG land use 1993: retail class corrected with urban expansion 1993 – 2008
- BBG land use 1996: retail class corrected with urban expansion 1993 – 2008
- BBG land use 2000: retail class corrected with urban expansion 2001 – 2008
- BBG land use 2003: retail class corrected with urban expansion 2001 – 2008
- BBG land use 2006: retail class corrected with retail in BBG land use map 2008
- BBG land use 2008 and onwards: no retail correction necessary

A point of attention is that this method depends on the land use area of retail and social-cultural services. Because the land use maps are always a generalized representation of reality, the area of retail locations and social-cultural services can be larger or smaller than in reality. Furthermore, in particular smaller locations are often omitted from the land use maps depending on the generalization rules applied. Although we can assume that larger retail areas attract in general more people and contribute to a larger walkability of an area, this relation is probably less obvious for several social-cultural services such as large hospital buildings, religious buildings or music halls. The size of these buildings is not always a good indicator for the number of walking people it attracts (e.g. the walking attraction of an hospital is centered mostly around its entrance. To prevent that these large land use areas contribute disproportional to the theoretical walkability of an area, we made an alternative weighted version of this component which can be used to test if this weighting gives better results in the subsequent validation process. In the weighting process we gave a weight of 5 to retail areas and a weight of 1 to social-cultural land use based on a rough guess that the floor space of many buildings with social-cultural services is up to 5 or more times bigger than the average retail location.

GIS operations for production of component maps and table variables

A summary of GIS operations carried out to produce the required table variables is presented here below. For exact procedures we refer to the Python scripts used to automate these GIS operations.

Adress/PC6 level

Python script: 2_Density_retail_and_service_destinations.py (first part)

- Correct retail class for land use maps 1989, 1993, 2000, 2003 and 2006 by erasing retail areas located in urban expansion areas that are more recent than the specific land use map (see correction scheme above, this part is not automated in Python)
- Convert vector to raster (25 meter cells)
- Reclassify / aggregate land use maps to suggested retail / services class (value = 1)
- Make weighted versions of reclassified land use maps (retail area * 5)
- Calculate focal statistics (summed area) per year and exposure area
- Use raster calculation to calculate retail / social services density for each gridcell. Resulting value is the area proportion or percentage of retail / services per 100 m² exposure area.
- Calculate z-score for each raster cell (raster cell value - Mean) / SD

Python script: 2_Retail_and_cultservice_extract_values_2_PC6.py

- Extract values to PC6 centroid table for each exposure variable

Geo-dms application with adapted lines in 'stam.dms'

- Extract values to address coordinate table for each exposure variable

PC4/Neighborhood level

Python script: 2_Density_retail_and_service_destinations.py (second part)

- Calculate aggregated statistics (sum / density) per neighbourhood and PC-4 area using the zonal statistics function (mean)

In Figure 4 a map example is given with (unweighted) Z-scores for the year 2015 within a 1000 meter walk range, while Figure 5 shows the weighted version of this map.

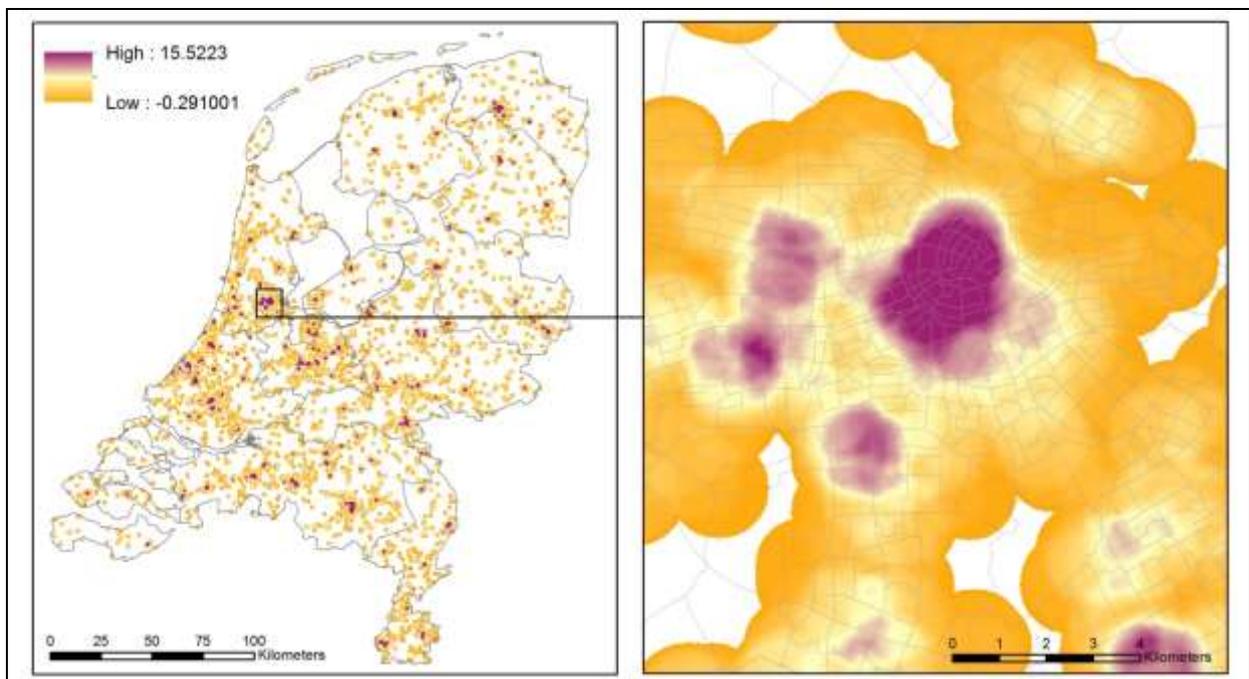


Figure 4 Map example: (unweighted) Z-scores for density of retail and service destinations 2015, 1000 meter radius

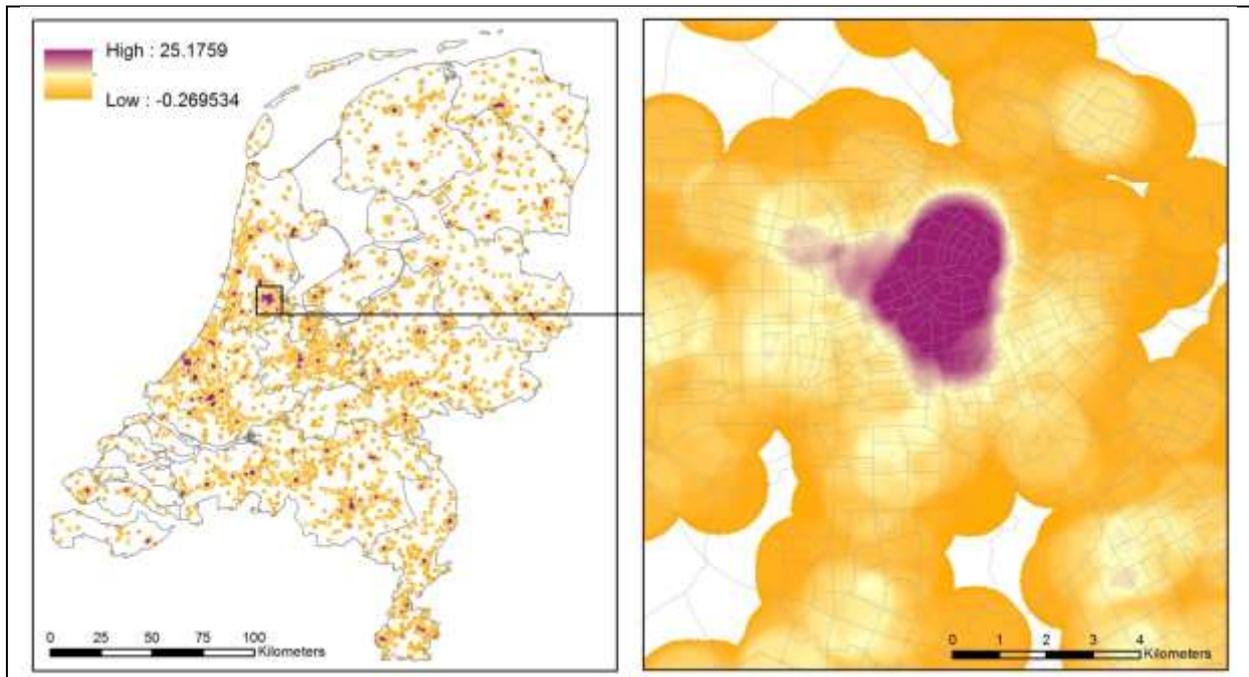


Figure 5 Map example: (weighted) Z-scores for density of retail and service destinations 2015, 1000 meter radius

Discussion

As already explained here above we have used the BBG land use map series to be able to produce comparable data for earlier years as well. Another advantage of these datasets is that they include the land use class 'Sociaal-culturele voorzieningen' as well. We also considered adding the class *public services* which contains public facilities such as police stations, city halls and court houses. However, we decided to leave the public services class out, since this class also contains many 'public' facilities that are less attractive as walking destinations, such as water treatment and waste incineration plants, storage areas, military areas (public according to CBS?) and utility companies.

The described correction procedure for the BBG maps prior to 2008 possibly excludes certain large shopping malls that were developed outside inhabited areas, which could be the case for the CBS map 'bevolkingskernen' for 2001. However, we think it is reasonable to assume that such large retail areas are already included in the BBG land use maps.

Another point of attention is that the current procedure relies on the relative spatial area of the different service destinations, while the area is not in all cases a good or reliable predictor of attraction for pedestrians. See e.g. the map example in Figure 6 in which the (unweighted) high densities (in purple) in the lower left part of the map are in reality a concentration of barracks and educational buildings.

We expect that the type of destination is more informative than its specific area size. In any case we expect in general a difference between retail destinations and social-cultural destinations which are often larger in area while not necessarily more attractive. Therefore, a second version of this data component was constructed with a weighting scheme applied (see previous section).



Figure 6 Land use map 2012 around Amersfoort (left) and (unweighted) density map (right) of retail and social-cultural services for 1000 meter radius walk range (gemeente Amersfoort)

The downside of choosing the BBG mapseries is that these data concern generalized data with unknown accuracy. The walkability definition document (Walkability working group, 2019) discusses different possible measures concerning the proximity, density and variety of retail options in a given area. It argues that *'proximity of retail'* which is associated with the number and variety of destinations within a specified distance of a given area, is a measure that could be constructed on the basis of LOCATUS, as it registers and geo-localises a variety of objectively measured retail branches (about 265 different branches).

Although there are some obvious advantages to this LOCATUS data, there are also some downsides to consider:

- Retail data are measured in a standardised way from 2004 onwards, not before (we need historical data from ca. 1985 onwards).
- High quality data, but not open data. VUmc has a datacontract with LOCATUS (which will probably be renewed this year), but with some limiting user conditions
- Potential other relevant facilities such as general practices and children day-care etc. are not included.

Other alternatives could be CBS data and/or BAG data. As indicated earlier in this document, with the 'vierkantstatistieken', CBS offers 100 and 500 meter resolution population data, including distance/density measures related to several relevant facilities (see Appendix 1). Operationalisation would be similar to that of population density (see text block in that paragraph). Unfortunately, these detailed data are only available from 2015 onwards (with 2016 and 2017 having progressively less completed data items).

BAG data can be useful as well as it gives for each building(unit) mapped the main user purpose, subdivided in meeting, detention, health care, industrial, education, sport, residence, store/retail and remaining functions. Although, BAG gives the construction year for each building, it only registers the current use of the building (same year as the dataset) and does not clearly registers changes in the

use of the building, even though it comes with a mutation field^v. Regarding the mutation field the following is stated (In Dutch):

(<https://www.amsterdam.nl/stelselpedia/bag-index/handboek-inwinnen/introductie-bag/mutaties-historie/>): *“In de BAG wordt de historie bewaard van alle objecten. Bij wijzigingen in deze objecten krijgt de oude situatie (zonder de wijziging) een eindGeldigheid en wordt een nieuw voorkomen van het betreffende object opgevoerd met het gewijzigde gegeven. Het attribuut [beginGeldigheid](#) van het nieuwe voorkomen is een datum, die dezelfde is als de eindGeldigheid van het oude voorkomen. Het voorkomen van een object met een eindGeldigheid die in het verleden ligt, is daarmee per definitie een historisch voorkomen van dat object.”*

Although it seems that it should be possible to generate relevant historical information using this field, this does not become directly clear when the data itself is accessed. This option should be looked at in more detail.

Summarizing, we could construct a time series from the BAG dataset but it possibly would have a questionable temporal accuracy. Main advantage of this dataset would be its spatial accuracy. A possibility could also be to use it in combination with one of the other datasets and use the building construction year as a hard way to filter out service destinations in specific data years.

Another option to consider is OpenStreetMap data which includes a point layer with many retail and many other service destinations. However, completeness and accuracy are unknown and also this data is not available (or less complete) for historical years. It may be worthwhile however to compare the quality of this data with e.g. Locatus data and construct rough historical versions of the dataset with the help of ‘pand-historie’ in the BAG data.

5. Land-use mix

Introduction

The third component in the composite walkability index is the so called ‘land-use mix’. Mixed or diverse land use is considered one of the “3Ds” —density, pedestrian friendly design, and diversity— that have been found to associate with walking (Brown et al., 2009). Mixed use brings many diverse walking destinations together in an area, which may be especially important for supporting walking for transportation purposes.

Land-use mix is commonly measured and accepted by means of an entropy index that indicates the heterogeneity of land use. This index is expressed as follows:

$$Entropy\ index = (-1) * \frac{[(\frac{b1}{a}) * \ln(\frac{b1}{a}) + (\frac{b2}{a}) * \ln(\frac{b2}{a}) + etc.]}{\ln(n)}$$

Where, *a* is the total area in square meter of the different land-uses, and *b* refers to a specific land-use category. Different land-use categories can be included, whereby *b1* refers to category one, *b2* to

^v <https://www.amsterdam.nl/stelselpedia/bag-index/handboek-inwinnen/introductie-bag/mutaties-historie/>

category two and so on. The variable n refers to the total number of land-use categories.^{vi} Four land-use categories are often used: 1) Industrial, commercial, public, military and private units, 2) Residential areas, 3) Green urban areas, and 4) Sports and leisure facilities.

Data selection and preparation

To build this index we use the land use map series BBG - 'Bestand BodemGebruik' which covers the Netherlands for the years 1989, 1993, 1996, 2000, 2003, 2006, 2008, 2010, 2012 and 2015 on a scale of 1:10,000 and geographically corresponds to the topographical map of the Netherlands (TOP10 NI / TOP10 vektor). Source locations are given in section 3 (Density of retail and service destinations). For the years 1989 to 2006 we used the version which was corrected for the class 'commercial' ('detailhandel' and 'horeca') as explained in section 3.

Land use classes change over the years but similar land use categories as suggested for the entropy index are available, such as the following built-up classes: Residential area, industrial area, Retail and catering, Public facilities, Public and socio-cultural facilities and open area classes like Parks and public gardens, sport parks, recreation areas and allotment gardens.

We decided to use the following 5 different land use classes or land use compositions:

Land use mix class	Classes from BBG
1. residential	woongebied en verblijfsrecreatie (recreatiewoningen)
2. commercial	detailhandel en horeca
3. social-cultural services	sociaal-culturele voorzieningen
4. offices and public services	kantoren en openbare voorzieningen
5. greenspace and recreation	parken en plantsoenen, bos, begraafplaats sportterreinen en dagrecreatieve terreinen

However, before these steps can be operationalized in GIS, we need to deal with a problematic class namely class number 4 industry and offices. This class is represented as the land use class business park ('bedrijfsterreinen') in the CBS bodemgebruikbestanden and concerns a mixed class that can represent industrial plants, business parks (trade / manufacturing, etc.) and office space. Following Brown et al. (2009), we assume that the presence of industrial terrains and business parks does not contribute to the walkability of a neighborhood, while on the contrary, offices can be considered walking destinations that contribute to utilitarian walking. This line of thought is also supported by Duncan et al. (2010) who demonstrate that significant relations between land use heterogeneity and walking behavior can only be expected when land uses with limited relevance for the type of walking studied are excluded. For this reason we decided to exclude industrial land use from our land-use mix and split up the class 'business park' to the classes 'business park and industrial plants' and the class 'office space' using the attribute field 'Gebruiksdoel' (= use objective) in the dataset Basisregistratie

^{vi} Bahadure, S. & Kotharkar, R. Assessing Sustainability of Mixed Use Neighbourhoods through Residents' Travel Behaviour and Perception: The Case of Nagpur, India. *Sustainability*. 2015.

Adressen en Gebouwen (BAG) for the disaggregation process. In this attribute field office functions are distinguished from functions for industrial processing/storage of materials or agricultural purposes. The BAG dataset we use concerns the year 2015. This means that for a certain land use mix year, e.g. 2012, it can happen that the use objective in BAG in 2015 is different from the use objective in 2012, but we assume the number of function changes between office space and business or industrial space will be relatively small.

Appendix 3 explains the technical details of the disaggregation process in GIS, which was not automated in Python.

GIS operations for production of component maps and table variables

A summary of GIS operations carried out to produce the required table variables is presented here below. For exact procedures we refer to the Python scripts used to automate these GIS operations.

Adress/PC6 level

Python script: 3_Land_use_mix_Walkability.py

- Convert vector to raster (25 meter cells)
- Reclassify land use maps to suggested classes (as far as possible)
- Calculate focal statistics (summed area) per year, combined land use classes and exposure area (component a)
- Calculate focal statistics (summed area) per year, land use class and exposure area (components b1, b2, b3 and b4)
- Use raster calculation to calculate entropy index for each gridcell. The resulting value is the (unitless) entropy index ranging from 0-1, with a higher score indicating a more heterogeneous land use mix
- Calculate z-score for each raster cell $(\text{raster cell value} - \text{Mean}) / \text{SD}$

Python script: 3_Land_use_mix_extract_values_2_PC6.py

- Extract values to PC6 centroid table for each exposure variable

Geo-dms application with adapted lines in 'stam.dms'

- Extract values to address coordinate table for each exposure variable

PC4/Neighborhood level

Python script: 3_Land_use_mix_24jul19_administrative_units.py

- Calculate zonal statistics per neighbourhood and PC-4 area and use raster calculation to calculate entropy index for neighbourhood / PC-4 area

In Figure 7 a map example is given with Z-scores for the year 2015 within a 1000 meter walk range.

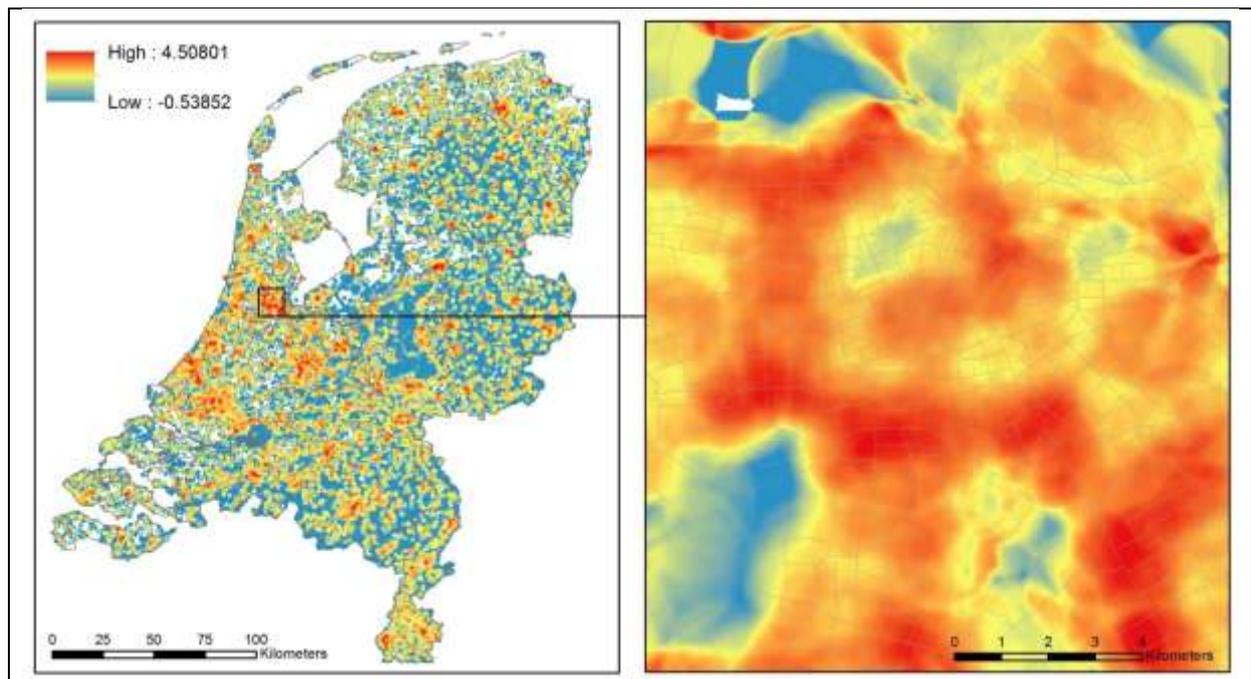


Figure 7 Map example: Z-scores for entropy index land use mix 2015, 1000 meter radius

Discussion

We considered to use a remaining class to prevent that the land use mix score becomes zero in e.g. a mainly forested area. However, adding all other land uses resulted in negative entropy scores and strictly speaking it makes sense that the diversity in the land use mix becomes zero when only one land use class is present. In the case of forest (green space) this will also be a separate component in the walkability score. Moreover, adding other classes would mean they will contribute to the heterogeneity of the land use while they are considered not relevant for peoples walking behavior (see further below). Therefore, we did not use a class with remaining land use categories.

Another point of concern is that according to Hajna et al. (2014) and also brought forward by Brown et al. (2009), the entropy index has a major limitation which should be addressed, because entropy remains constant when distinct land-use types remain in constant relative proportions; yet mixing or integration of land uses can change dramatically. For example for the entropy index it does not make any difference if the proportions between residential and commercial land use are 80/20 or 20/80, while this makes a big difference for walking behavior, e.g. for the amount of transport walking. Im and Choi (2018) made this effect clear in a figure, see Figure 8, and named it the 'n-shaped characteristic' of the LUM (Land-Use Mix) index, but do not offer other alternatives for the LUM index than recommendations for its proper use.

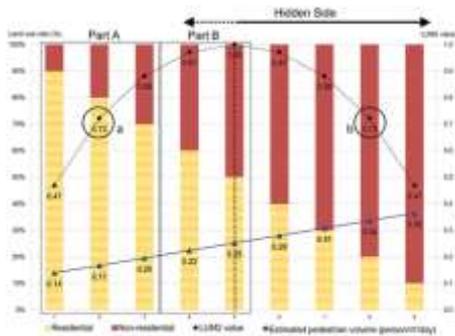


Figure 8 The 'n-shaped characteristic' of the LUM (Land-Use Mix) index according to Im and Choi (2018)

If we inspect our resulting land use mix maps, we see indeed different unexpected patterns that relate probably to the problem described here above. In the map example in Figure 9 that displays the land use mix entropy for a walk range of 1000 meter radius for the municipality of Amersfoort, we see among others high unexpected entropy values in the northwest part of the map (indicated with the blue circle) around an area dominated by industry and company sites. These effects are particularly visible in the maps based on large buffer radii and therefore we recommend to improve this component and maybe avoid to use buffer sizes larger than e.g. 750 meter.



Figure 9 Map example entropy index land use mix, with 1000 meter radius walk range (municipality of Amersfoort)

6. Street connectivity (intersection density)

Introduction

The fourth component for the walkability index is the street connectivity or intersection density for which frequently positive associations have been found with physical activity levels (see e.g. Grasser et al., 2013), which measures the connectivity of the street network, represented by the ratio between the number of true intersections (three or more legs) to the size of the selected area. The idea behind the use of street connectivity is that in general higher density of intersections corresponds with a more direct (and therefore shorter) path between destinations.

Data selection and preparation

An important data source to calculate this measure is the TOP10 NL and TOP10 vektor, the topographical map of the Netherlands, scale 1:10.000 available for 2003, 2005^{vii}, 2011, 2012, 2015 and 2019.

Source data for the years 2003, 2005, 2011 and 2012 were acquired from DANS-KNAW (Data Archiving and Networked Services). Preprocessed source data for the years 2015 and 2019 was downloaded from the data service of Esri Nederland (via organisation account).

The datasets contain a specific point file named 'Wegdeel_hartpunten' which contains information on types of intersection and connection roads. It is important to reclassify this data to different road types and/or transport users prior to summarizing numbers of intersections per area unit. Relevant classes are 'Hoofdverkeersgebruik' and 'Hoofdverkeersgebruik_CSV', see table below.

Class from Hoofdverkeersgebruik	Class from Hoofdverkeersgebruik_CSV
-gemengd verkeer	Gemengd verkeer Busverkeer fietsers, bromfietsers voetgangers
-voetgangers	Busverkeer gemengd verkeer fietsers, bromfietsers voetgangers Voetgangers overig Voetgangers
-snelverkeer	Busverkeer gemengd verkeer voetgangers
-busverkeer	Busverkeer gemengd verkeer fietsers, bromfietsers voetgangers
-vliegverkeer	Gemengd verkeer fietsers, bromfietsers voetgangers
-fietsers, bromfietsers	Gemengd verkeer voetgangers
-overig	Overig

We have used the following attribute selection procedure in GIS -not automated in Python- from the main class 'Hoofdverkeersgebruik'.

SELECT FROM WEGDEEL_HARTPUNT WHERE:

```
HOOFDVERKEERSGEBRUIK_CSV = 'busverkeer|fietsers, bromfietsers|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'busverkeer|gemengd verkeer|fietsers, bromfietsers|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'busverkeer|gemengd verkeer|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'busverkeer|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'fietsers, bromfietsers|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'gemengd verkeer|fietsers, bromfietsers|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'gemengd verkeer|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'snelverkeer|fietsers, bromfietsers|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'snelverkeer|gemengd verkeer|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'snelverkeer|voetgangers' OR HOOFDVERKEERSGEBRUIK_CSV = 'voetgangers'
```

Result (TOP10 2015): 171.113 of 1.121.934 selected

Add to selection:

^{vii} Only per individual mapsheet in dxf-format (!)

HOOFDVERKEERSGEBRUIK_1 = 'gemengd verkeer' OR HOOFDVERKEERSGEBRUIK_1 = 'overig' OR
HOOFDVERKEERSGEBRUIK_1 = 'voetgangers'

Result (TOP10 2015): 1.046.309 of 1.121.934 selected

Remove from selection (2015 and earlier):

"TYPEWEG_1" = 'autosnelweg' OR "TYPEWEG_1" = 'hoofdweg' OR "VERHARDI_1" = '> 7 meter'

For 2019:

"TYPEWEG_1" = 'autosnelweg' OR "TYPEWEG_1" = 'hoofdweg' OR VERHARDINGSBREEDTEKLASSE = '> 7 meter'

To produce also maps for historical situations before 2003 for which we do not have topographical maps on 1:10.000 scale available, we have used the CBS 'bevolingskernen' map of 2001 and the PBL urbanization maps ('stedelijke gebieden') of 1989 and 1993 to remove street connections that were not yet existing in those periods.

GIS operations for production of component maps and table variables

A summary of GIS operations carried out to produce the required table variables is presented here below. For exact procedures we refer to the Python scripts used to automate these GIS operations.

Adress/PC6 level

- Make attribute query on WEGDEEL_HARTPUNT layer to select relevant points for walking (see above)
- Export to separate layer with selected 'hartpunten'
- Remove pedestrian areas to create sidewalk maps for 1989, 1993 and 2001
- The selection of hartpunten 2003 is carried out via a 'selection by location' based on intersection of hartpunten 2012 and TOP10 roadlayers 'vlak 200' and 'vlak 300' (polygon layers containing all road surfaces, as the TOP10 of 2003, does not contain a separate hartpunten road intersection layer)

Python script: 4_Street-connectivity.py (first part)

- Convert vector to raster (25 meter cells)
- Calculate focal statistics per year (sum)
- The resulting value is the number of road connections per hectare (including roads suitable for pedestrians)
- Calculate z-score for each raster cell (raster cell value - Mean) / SD)

Python script: 4_Street-connectivity_extract_values_2_PC6.py

- Extract values to PC6 centroid table for each exposure variable

Geo-dms application with adapted lines in 'stam.dms'

- Extract values to address coordinate table for each exposure variable

PC4/Neighborhood level

Python script: 4_Street-connectivity.py (second part)

- Calculate aggregated statistics (sum / density) per neighbourhood and PC-4 area

In Figure 10 a map example is given with Z-scores for the year 2015 within a 1000 meter walk range.

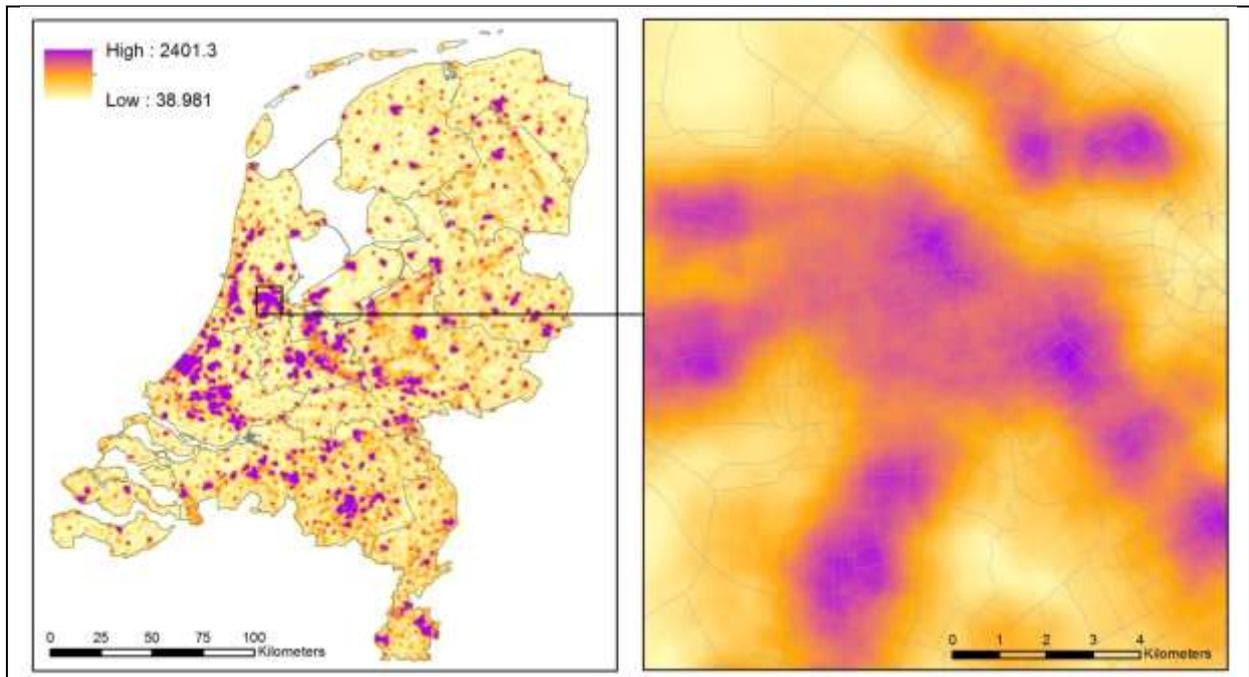


Figure 10 Map example: Z-scores for street intersection density 2015, 1000 meter radius

Discussion

In general we see that large transport nodes such as roundabouts and street crossings give the highest connectivity values, which can be representative for the traffic as a whole, but less for pedestrians. It is therefore important to keep larger car roads as much as possible out of the selections. An option could be, to select only the roads that border on sidewalk areas (see section 7).

Another point of attention concerns the size of the walk ranges. Compare for example the maps below in Figure 11 that show the connectivity of a part of Amsterdam for an walk range of 500 meter (left) and 1000 meter (right), where the 1000 meter map shows very little variation in connectivity within the city limits.

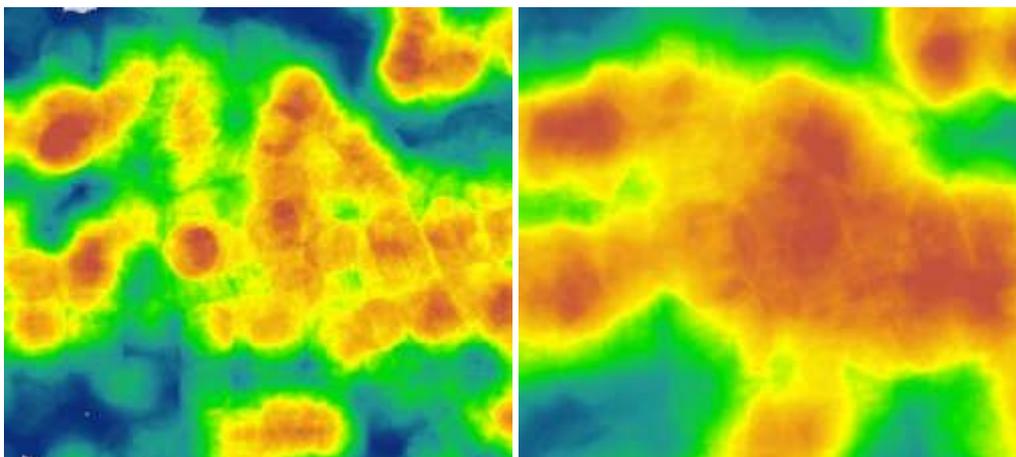


Figure 11 Street intersection density 2012 Amsterdam South of 'het IJ', 500 meter (left), 1000 meter (right)

7. Green space density

Introduction

Green space and urban greenness is especially considered important for recreational types of walking (see e.g. Sugiyama et al., 2013 and Vich et al., 2019) and was therefore added as an additional component to the walkability score.

Data selection and preparation

We define green space as a combination of the following land use classes from land use map series BBG - 'Bestand BodemGebruik' (years 1989, 1993, 1996, 2000, 2003, 2006, 2008, 2010, 2012, 2015):

- Parken en plantsoenen
- Bos
- Begraafplaats

Source locations are given in section 3 (Density of retail and service destinations).

GIS operations for production of component maps and table variables

A summary of GIS operations carried out to produce the required table variables is presented here below. For exact procedures we refer to the Python scripts used to automate these GIS operations.

Adress/PC6 level

Python script: 5_Green_space.py (first part)

- Convert vector to raster (25 meter cells)
- Reclassify / aggregate land use maps to suggested green space class (value = 1)
- Calculate focal statistics (summed area) per year and exposure area
- Use raster calculation to calculate green space density for each gridcell. Resulting value is the area proportion or percentage of green space per 100 m² exposure area
- Calculate z-score for each raster cell (raster cell value - Mean) / SD

Python script: 5_Green_space_extract_values_2_PC6.py

- Extract values to PC6 centroid table for each exposure variable

Geo-dms application with adapted lines in 'stam.dms'

- Extract values to address coordinate table for each exposure variable

PC4/Neighborhood level

Python script: 5_Green_space.py (second part)

- Calculate aggregated statistics (sum / density) per neighbourhood and PC-4 area

In Figure 12 a map example is given with Z-scores for the year 2015 within a 1000 meter walk range.

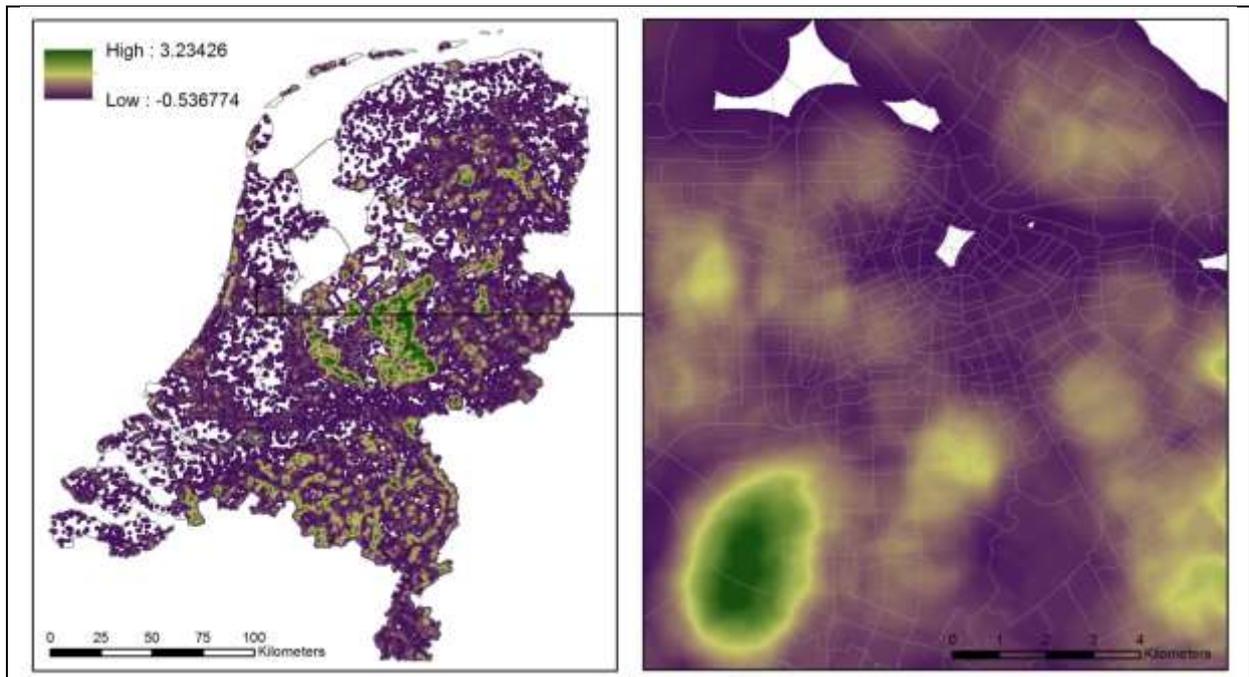


Figure 12 Map example: Z-scores for green space density 2015, 1000 meter radius

Discussion

Green space is not a common component in walkability indices but we have added it as a separate component for its expected effect on recreational walking.

8. Side walk density

Introduction

Literature shows that neighborhood sidewalk length is positively associated with *participation* and minutes of neighborhood-based walking for transportation, also when corrected for residential self-selection (McCormack et al., 2012, Veerman et al., 2016). However, no association was found for neighborhood-based walking for recreation and other studies show that although the presence of sidewalks increases the likelihood of walking it has no apparent effect on the frequency of walks taken.

In the Basisregistratie Grootchalige Topografie (BGT) sidewalks, sidewalk areas, pedestrian stairs and residential areas suitable for walking (“woonerven”) are mapped as polygons on a very accurate 1:500 to 1:5000 scale. Assuming that sidewalk lengths correlate well to sidewalk areas we use the walkable map features from this dataset for the construction of this component.

Data selection and preparation

The daily actualized BGT can be downloaded from: <https://mijn.pdok.nl/nl/producten/pdok-downloads/download-basisregistratie-grootchalige-topografie>. This concerns however a zipped

collection of around 80 separate GML files, which requires a lot of processing to get the desired format. We have used therefore fileGDB version of this dataset (2019) which is provided by ESRI-NI via the ArcGIS online website / ESRI-NI user group (account needed): 'Datasets - Esri Nederland'.

There are no historical BGT files available but we can roughly produce sidewalk areas of earlier years by removing sidewalk cells located in areas that were not yet developed into residential or infrastructural land uses. We use the building and road features from the TOP10 2003, 2012 and 2015 to produce the sidewalk areas for the corresponding years. We use the land use maps of the Netherlands (BBG) of 1989, 1996, 2000 and 2008 (in combination with TOP10) to produce the remaining corresponding years. To remove sidewalks that we consider 'non existing' before 2019, we carry out a 'Select by location' procedure in ArcGIS in which we select and remove sidewalks that do not intersect with either roads or buildings or woongebied in TOP10 (2003, 2012 and 2015) or in BBG (land use classes 11, 12, 20, 21, 22, 23, 24, 32, 40, 41, 42, 43, 44 and 60 for land use maps 1996, 2000, 2003 and 2008 and classes 21, 31, 33, 34, 35, 36, 42, 43, 44, 51, 52, 53, 54, 55 and 73 for land use maps 1989 and 1993). In the spatial selection method we apply a small searching distance of 10 meters around the roads and buildings to make sure intersection with present sidewalks will take place and we assume that the majority of sidewalks is located directly alongside buildings and other roads. An exception could be the sidewalks that are located e.g. in forests, parks and natural areas. A number of such sidewalks could be included in the historical maps, while in reality they might not (yet) be present.

From the BGT 2019 we select the vector layer 'wegdeel_v' and use the field 'functie' to select all 'voetpaden', 'voetgangersgebied', 'voetpad op trap' (2.169.921 records), and export the selected polygons and subsequently convert to a grid with a resolution of 5x5 meter in which rastercells have the value '1' if they represent sidewalk area.

Next, we calculate the neighbourhood statistics (SUM) for each walk range between 150 and 2000 meter. This gives us the number of sidewalk cells in each walk range. This 5 meter grid (which is very large) we resample to a 25 meter grid after which we calculate the percentage sidewalk area per walk range.

GIS operations for production of component maps and table variables

A summary of GIS operations carried out to produce the required table variables is presented here below. For exact procedures we refer to the Python scripts used to automate these GIS operations.

Adress/PC6 level

- Select vektorlayer wegdeel_v
- Select 'voetpad', 'voetgangersgebied', 'voetpad op trap' from field 'functie'

Python script: 6_Sidewalk_density.py (first part)

- Export selection to polygon layer and convert vector to raster (5 meter cells) with value 1 for sidewalks
- Take focal SUM of SidewalksBGT2012_5m (value = 1) for each walk range radius (neighbourhood size)
- Resample the sidewalk 5 meter grid to a 25 meter grid, using the bilinear sampling method
- Calculate % sidewalk areas per walk range radius by using raster calculation:
- $((\text{resamp25m_focstatsidewalk_2_5}) * 25) / \pi * \text{square}(r) * 100$
- Calculate z-score for each raster cell $(\text{raster cell value} - \text{Mean}) / \text{SD}$

Python script: 6_Sidewalk_density_extract_values_2_PC6.py

- Extract values to PC6 centroid table for each exposure variable

Geo-dms application with adapted lines in 'stam.dms'

- Extract values to address coordinate table for each exposure variable

PC4/Neighborhood level

Python script: 6_Sidewalk_density.py (second part)

- Calculate aggregated statistics (sum / density) per neighbourhood and PC-4 area

In Figure 13 a map example is given with Z-scores for the year 2015 within a 1000 meter walk range.

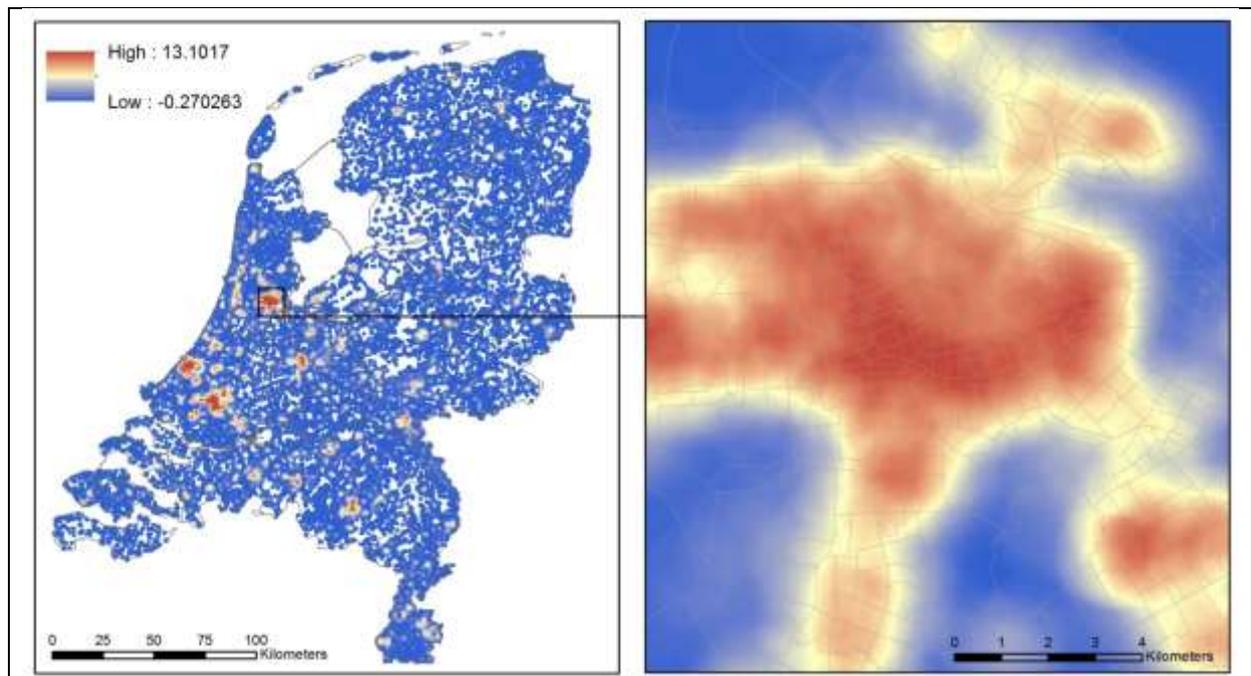


Figure 13 Map example: Z-scores for side walk density 2015, 1000 meter radius

Discussion

9. Walkability index

The final result of the GIS operations is a set of tables in which for each residential address location in the Netherlands an attribute value is available for each of the six walkability components. All component values are converted in z-scores (representing the number of standard deviations from the population mean) before summing them, according to a weighting scheme (to be decided) in accordance with a previously developed walkability index or other expert opinion. The combination and computation into a total walkability score for each selected year, each location and buffer size is

carried out within SPSS using the following formula:

Walkability = (z-population density) + (z-street connectivity) + (z-land use mix) + (z-retail environment) + (z-green space) + (z-pedestrian area density)

Finally, the score is normalized to derive values for the walkability index between 0 and 100. The higher the score, the greater the ability to walk in that area. Based on the XY, PC6, PC4 locations or neighborhoods of participants in the cohort data the walkability index score for each raster cell in the Netherlands can be extracted as an additional variable to each record in the cohort participant table after which the validation process can begin.

10. Discussion

We have based our current index on available scientific literature about walkability (in combination with the possibilities offered by the available geodata), which is dominated by studies from areas such as the United States where factors that determine walkability may radically differ from European areas and the Netherlands in particular. It makes therefore sense to look critically to the components and criteria used and evaluate the application to specific European and Dutch settings.

For the current version of the walkability score we have applied circular walk ranges around each gridcell location, based on straight-line Euclidean distance. However, from a walkability standpoint this is not an optimal method because it doesn't account for physical barriers such as building blocks, water bodies, railway lines and so on. A much better approach would therefore be the implementation of Service Areas, which means the area that can be reached by foot using the walkable infrastructure network. This involves however the calculation of more than 50 million service areas of irregular shapes for each separate transport mode. More problematic still would be the subsequent step calculating summary statistics for each particular service area for each component of the walkability index. Such an approach would be feasible in the case of a limited number of known residential locations, but not in our case where we have to calculate a score for each and every location (on a 25 meter resolution) for the whole of the Netherlands.

Population density

Density of retail and service destinations (retail environment)

Land-use mix

Street connectivity (intersection density)

Green space

Side walk presence

11. References

- Brown, B. B., Yamada, I., Smith, K. R., Zick, C. D., Kowaleski-Jones, L., & Fan, J. X. (2009). Mixed land use and walkability: Variations in land use measures and relationships with BMI, overweight, and obesity. *Health & place, 15*(4), 1130-1141.
- Duncan, M. J., Winkler, E., Sugiyama, T., Cerin, E., duToit, L., Leslie, E., & Owen, N. (2010). Relationships of land use mix with walking for transport: do land uses and geographical scale matter? *Journal of urban health : bulletin of the New York Academy of Medicine, 87*(5), 782–795.
- Frank, L. D., Sallis, J. F., Saelens, B. E., Leary, L., Cain, K., Conway, T. L., & Hess, P. M. (2010). The development of a walkability index: application to the Neighborhood Quality of Life Study. *British journal of sports medicine, 44*(13), 924-933.
- Grasser, G., Van Dyck, D., Titze, S., & Stronegger, W. (2013). Objectively measured walkability and active transport and weight-related outcomes in adults: a systematic review. *International journal of public health, 58*(4), 615-625.
- Hajna, S., Dasgupta, K., Joseph, L., & Ross, N. A. (2014). A call for caution and transparency in the calculation of land use mix: measurement bias in the estimation of associations between land use mix and physical activity. *Health & place, 29*, 79-83.
- Im, H. N., & Choi, C. G. (2018). The hidden side of the entropy-based land-use mix index: Clarifying the relationship between pedestrian volume and land-use mix. *Urban Studies, 0042098018763319*.
- Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. (2007). Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. *Health & place, 13*(1), 111-122.
- McCormack, G. R., Shiell, A., Giles-Corti, B., Begg, S., Veerman, J. L., Geelhoed, E., ... & Emery, J. H. (2012). The association between sidewalk length and walking for different purposes in established neighborhoods. *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 92.
- Shashank, A., & Schuurman, N. (2019). Unpacking walkability indices and their inherent assumptions. *Health & place, 55*, 145-154.
- Sugiyama, T., Neuhaus, M., Cole, R., Giles-Corti, B., & Owen, N. (2012). Destination and route attributes associated with adults' walking: a review. *Medicine and science in sports and exercise, 44*(7), 1275-1286.
- Sugiyama, T., Giles-Corti, B., Summers, J., du Toit, L., Leslie, E., & Owen, N. (2013). Initiating and maintaining recreational walking: a longitudinal study on the influence of neighborhood green space. *Preventive medicine, 57*(3), 178-182.
- Veerman, J. L., Zapata-Diomed, B., Gunn, L., McCormack, G. R., Cobiac, L. J., Herrera, A. M. M., ... & Shiell, A. (2016). Cost-effectiveness of investing in sidewalks as a means of increasing physical activity: a RESIDE modelling study. *BMJ open, 6*(9), e011617.

Vich, G., Marquet, O., & Miralles-Guasch, C. (2019). Green streetscape and walking: exploring active mobility patterns in dense and compact cities. *Journal of Transport & Health, 12*, 50-59.

Walkability group (2019). Defining and applying measures of walkability. Internal document by group members VU University Medical Center, Amsterdam UMC) and the Global and Geo Health Data Centre (GGHDC), Utrecht University.

Appendix 1 Example attribute list of the 'vierkantstatistieken'

Name	Value
Aantal_inwoners	295
Aantal_mannen	150
Aantal_vrouwen	145
Aantal_inwoners_0_tot_15_jaar	20
Aantal_inwoners_15_tot_25_jaar	75
Aantal_inwoners_25_tot_45_jaar	145
Aantal_inwoners_45_tot_65_jaar	45
Aantal_inwoners_65_jaar_en_ouder	10
Percentage_autochtonen	60
Percentage_westerse_allochtonen	20
Percentage_niet_westerse_allochtonen	20
Aantal_part_huishoudens	210
Aantal_eenpersoonshuishoudens	140
Aantal_meerpersoonshuishoudens_zonder_kind	45
Aantal_eenouderhuishoudens	5
Aantal_tweeouderhuishoudens	10
Gemiddelde_huishoudensgrootte	1.4
Aantal_woningen	180
Aantal_woningen_bouwjaar_voor_1945	135
Aantal_woningen_bouwjaar_45_tot_65	-99997
Aantal_woningen_bouwjaar_65_tot_75	35
Aantal_woningen_bouwjaar_75_tot_85	-99997
Aantal_woningen_bouwjaar_85_tot_95	-99997
Aantal_woningen_bouwjaar_95_tot_05	-99997
Aantal_woningen_bouwjaar_05_tot_15	10
Percentage_koopwoningen	20
Percentage_huurwoningen	80
Aantal_huurwoningen_in_bezit_woningcorporaties	20
Aantal_meergezins_woningen	175
Aantal_leegstaande_woningen	15
Aantal_personen_met_uitkering_zonder_aow	25
Grote_supermarkt_gemiddelde_afstand_in_km	0.1
Grote_supermarkt_aantal_binnen_1_km	12.4
Grote_supermarkt_aantal_binnen_3_km	79.9
Grote_supermarkt_aantal_binnen_5_km	141.6
Winkels_ov_dagelijkse_levensm_gem_afst_in_km	0.0
Winkels_ov_dagel_levensm_aantal_binnen_1_km	98.4
Winkels_ov_dagel_levensm_aantal_binnen_3_km	631.2
Winkels_ov_dagel_levensm_aantal_binnen_5_km	1020.9
Warenhuis_gemiddelde_afstand_in_km	0.1
Warenhuis_aantal_binnen_5_km	15.0
Warenhuis_aantal_binnen_10_km	26.0
Warenhuis_aantal_binnen_20_km	41.1
Cafe_gemiddelde_afstand_in_km	0.2
Cafe_aantal_binnen_1_km	71.3
Cafe_aantal_binnen_3_km	603.7
Cafe_aantal_binnen_5_km	813.5
Cafetaria_gemiddelde_afstand_in_km	0.0
Cafetaria_aantal_binnen_1_km	78.2
Cafetaria_aantal_binnen_3_km	588.3
Cafetaria_aantal_binnen_5_km	847.2
Hotel_gemiddelde_afstand_in_km	0.3
Hotel_aantal_binnen_5_km	352.0
Hotel_aantal_binnen_10_km	391.0
Hotel_aantal_binnen_20_km	451.8
Restaurant_gemiddelde_afstand_in_km	0.0
Restaurant_aantal_binnen_1_km	189.6
Restaurant_aantal_binnen_3_km	1273.8
Restaurant_aantal_binnen_5_km	1714.9
Buitenschoolse_opvang_gem_afstand_in_km	0.3
Buitenschoolse_opvang_aantal_binnen_1_km	10.4
Buitenschoolse_opvang_aantal_binnen_3_km	64.6

Buitenschoolse_opvang_aantal_binnen_5_km	147.1
Kinderdagverblijf_gemiddelde_afstand_in_km	0.3
Kinderdagverblijf_aantal_binnen_1_km	24.2
Kinderdagverblijf_aantal_binnen_3_km	138.5
Kinderdagverblijf_aantal_binnen_5_km	251.2
Brandweerkazerne_gemiddelde_afstand_in_km	0.8
Oprit_hoofdverkeersweg_gemiddelde_afstand_in_km	1.9
Overstapstation_gemiddelde_afstand_in_km	3.0
Treinstation_gemiddelde_afstand_in_km	2.0
Attractiepark_gemiddelde_afstand_in_km	2.4
Attractiepark_aantal_binnen_10_km	4.0
Attractiepark_aantal_binnen_20_km	8.0
Attractiepark_aantal_binnen_50_km	47.0
Bioscoop_gemiddelde_afstand_in_km	0.4
Bioscoop_aantal_binnen_5_km	12.0
Bioscoop_aantal_binnen_10_km	13.0
Bioscoop_aantal_binnen_20_km	18.0
Museum_gemiddelde_afstand_in_km	1.0
Museum_gemiddeld_aantal_binnen_5_km	42.0
Museum_gemiddeld_aantal_binnen_10_km	48.0
Museum_gemiddeld_aantal_binnen_20_km	66.0
Theater_gemiddelde_afstand_in_km	0.9
Theater_aantal_binnen_5_km	36.0
Theater_aantal_binnen_10_km	49.0
Theater_aantal_binnen_20_km	61.0
Bibliotheek_gemiddelde_afstand_in_km	0.7
Kunstijsbaan_gemiddelde_afstand_in_km	4.7
Poppodium_gemiddelde_afstand_in_km	1.2
Sauna_gemiddelde_afstand_in_km	0.5
Zonnebank_gemiddelde_afstand_in_km	0.3
Zwembad_gemiddelde_afstand_in_km	0.8
Basisonderwijs_gemiddelde_afstand_in_km	0.3
Basisonderwijs_aantal_binnen_1_km	5.3
Basisonderwijs_aantal_binnen_3_km	49.7
Basisonderwijs_aantal_binnen_5_km	106.8
Havo_vwo_gemiddelde_afstand_in_km	0.5
Havo_vwo_aantal_binnen_3_km	14.3
Havo_vwo_aantal_binnen_5_km	23.0
Havo_vwo_aantal_binnen_10_km	37.0
Vmbo_gemiddelde_afstand_in_km	0.3
Vmbo_aantal_binnen_3_km	11.3
Vmbo_aantal_binnen_5_km	24.1
Vmbo_aantal_binnen_10_km	44.0
Voortgezet_onderwijs_gem_afstand_in_km	0.3
Voortgezet_onderwijs_aantal_binnen_3_km	23.3
Voortgezet_onderwijs_aantal_binnen_5_km	41.1
Voortgezet_onderwijs_aantal_binnen_10_km	67.0
Huisartsenpraktijk_gemiddelde_afstand_in_km	0.3
Huisartsenpraktijk_aantal_binnen_1_km	12.6
Huisartsenpraktijk_aantal_binnen_3_km	72.0
Huisartsenpraktijk_aantal_binnen_5_km	135.3
Ziekenhuis_excl_buitenpolikliniek_gem_afst_in_km	2.2
Ziekenhuis_excl_buitenpoli_aantal_binnen_5_km	2.0
Ziekenhuis_excl_buitenpoli_aantal_binnen_10_km	8.0
Ziekenhuis_excl_buitenpoli_aantal_binnen_20_km	11.0
Ziekenhuis_incl_buitenpolikliniek_gem_afst_in_km	2.2
Ziekenhuis_incl_buitenpoli_aantal_binnen_5_km	3.0
Ziekenhuis_incl_buitenpoli_aantal_binnen_10_km	10.0
Ziekenhuis_incl_buitenpoli_aantal_binnen_20_km	15.0
Apotheek_gemiddelde_afstand_in_km	0.1

Appendix 2. Development procedure for component residential density

This appendix describes the possible development of the geocomponent residential density as an alternative to population density, also sometimes used in the literature.

Residential density is described in the walkability document as the number of residential units per km² within a neighbourhood. It is sometimes calculated as the ratio of the number of dwelling units to the land area for residential use within each area (administrative neighbourhood, or buffer(s)). For this latter measure, data would be needed on number and location of houses/apartments, as well as the surface of land devoted for residential use.

Advantage of this measure is that we can base it on the spatially very accurate data of the Basisregistratie Adressen en Gebouwen (BAG), which maps every residential building and living unit/house hold (thus also separate stores in multi-store buildings) in the Netherlands from the year 2015 onwards [*to be checked*], see map example below.



Zoomed in map view of BAG for part of neighbourhood 'De Pijp', in Amsterdam showing attribute Table of multistore residential building

This means we can map the residential density per km² very accurately for recent years (2018, 2017, 2016 and 2015). Disadvantage of this dataset, is that the residential density is not always comparable to population densities and these figures are not available in this detail for historical years.

Both residential density and residential 'ratio' can be calculated from the BAG or from a combination of the BAG with a land use dataset indicating residential use (this could be derived from the BBG data - Bestand BodemGebruik or more precisely from the TOP10 vektor).

Operationalisation in GIS (BAG):

- Select residential objects in point file 'verblijfsobjecten'
- Carry out spatial join between residential objects (points) and buildings (polygons) and add column to attribute table indicating the number of vertical living unites per building
- Convert vector to raster (100 meter raster cells with number of living units as gridcell value)
- Calculate neighbourhood statistics (focal mean) for different neighbourhood radii
- Extract values to PC6 table for each exposure variable

Appendix 3. Steps for disaggregation land use class business park CBS Bodemgebruik

The following steps were carried out to disaggregate the land use class business park CBS Bodemgebruik to classes 'business park and industrial plants' and the class 'office space' using attribute field 'Gebruiksdoel' in BAG.

1. Select points with Gebruiksdoel1 = Industriefunctie from BAG point layer 'verblijfsobject' (other mixed function groups were not selected)
2. Use (1) to Select by location units from the BAG layer 'Pand' that intersect with 1.
3. Export selected polygons to new layer named 'Industry2015_BAG_BBG2012.shp'. Add field named 'Industrie'
4. Repeat 1 to 3 for Gebruiksdoel1 = Kantoorfunctie and export to layer 'Kantoor2015_BAG_BBG2012.shp'. Add field named 'Kantoor'
5. Select class 'bedrijfsterreinen' from BBG and export to separate polygon layer named 'Bedrijfsterreinen_2012.shp'.
6. Carry out spatial join (intersect) between 'Bedrijfsterreinen_2012' (target layer) and 'Kantoor2015_BAG_BBG2012.shp' (join layer). Make sure you set a merge rule on the field 'SHAPE_AR_1': SUM. This takes care all that the areas in m2 of all the kantoor units that are joined to a certain polygon in Bedrijfsterreinen_2012 are added to a total area per polygon. Name result: 'BBG2012_SpatJoinBAG2015_kantoor.shp'
7. Next carry out spatial join (intersect) between 'BBG2012_SpatJoinBAG2015_kantoor.shp' (target layer) and 'Industry2015_BAG_BBG2012.shp' (join layer). Make sure you set a merge rule on the field 'SHAPE_AR_2': SUM. This takes care all that the areas in m2 of all the industry units that are joined to a certain polygon in BBG2012_SpatJoinBAG2015_kantoor.shp are added to a total area per polygon. Name result: 'BBG2012_SpatJoinBAG2015_kantoor_industry.shp'.
8. Now add a new field (integer short) to 'BBG2012_SpatJoinBAG2015_kantoor_industry.shp' named 'Type_new'. Carry out the following attribute selections:
SELECT FROM 'BBG2012_SpatJoinBAG2015_kantoor_industry.shp' WHERE:
"SHAPE_Ar_1" > "SHAPE_Ar_2" (total office space is larger than total industrial space). Now use the Field calculator to give a value of 1 to the selected cells.
SELECT FROM 'BBG2012_SpatJoinBAG2015_kantoor_industry.shp' WHERE:
"SHAPE_Ar_2" >= "SHAPE_Ar_1" (total industrial space is larger than total office space). Now use the Field calculator to give a value of 2 to the selected cells.
SELECT FROM 'BBG2012_SpatJoinBAG2015_kantoor_industry.shp' WHERE:
"SHAPE_Ar_1" = "SHAPE_Ar_2" (total industrial space is equal to total office space, in most cases these concern polygon without values (0 m2) for office and industrial space). Now use the Field calculator to give a value of 0 to the selected cells.
9. Export only the new field 'Type_new' to a separate layer (name it e.g. 'Type_new2012.shp') and select only 'Kantoor polygons' and use this layer to carry out a spatial selection (Select by location / 'Are identical to the source layer feature') in BBG2012. Add a new field to the BBG2012 layer named 'BG2012Updt' and assign the value 25 for office space to the selected records. Switch the selection and give all other records their original value from the field 'BG2012A'.

Appendix 4. Planned or processed improvements walkability score

Component	Date	Improvement	Status
Population density	10 September 2019	Some (or all?) of the downloaded neighborhood maps prior to 2015 appear to contain irregularities (e.g. duplicate records, wrong numbers for areas and population densities) resulting from join errors with duplicate neighborhood names. These need to be checked and corrected. Join complete tables to neighborhood maps to prevent errors	To be done
Density of retail and service destinations			
Land-use mix			
Street connectivity (intersection density)	29 August 2019	Make spatial selection from WEGDEEL_HARTPUNT based on intersection with voetpaden_2019 (within search distance of 15 meter): 687166 records Add HOOFDVERKEERSGEBRUIK_1 = 'voetgangers': 712.976 records	In progress
Green space	13 September 2019	Replace (or add) CBS park and forest layer for bomenkaart Nederland (2017) for a more accurate 2016/2017 version of the score	Idea
Side walk presence	29 August 2019	Add 'Woonerf' to dataset from BGT Calculate sidewalk length instead of sidewalk area (dissolve polygons => close gaps < 100 m2 => convert poly to line => collapse dual line to centre line => convert polyline to raster with line-length) Calculate exact polygon area per 25 meter cell	In progress

-