Enabling a national road and street database in population statistics: Commuting distances for all employed persons and other accessibility statistics

Pasi PIELA

Statistics Finland, Finland

Abstract

Accessibility research is a highly relevant part of today's geographic information science and systems (GIS). Enabling a national road and street database, "Digiroad" in Finland, contributes to meeting new user needs in development of official statistics. It gives access to better information on service areas and distances to actual services such as public schools, or even to upgrade commuting statistics on distances between each home and work point coordinates along the actual road network.

As one case example, commuting distances for the employed Finnish population is given in the paper. The actual pairwise computations have been made for all 2.1 million employed persons that have coordinate pairs. Naturally, the target is to estimate the shortest paths. Besides final results from these, technical challenges are also addressed in the paper.

However, travel time is even a hotter topic in accessibility. Another case example is given on emergency coverage based on drive time.

Key Words: accessibility, route optimisation, pairwise distances, travel time.

1. Data Warehouse

Digiroad is a national database managed by the Finnish Transport Agency, which contains accurate data on the location of all roads and streets in Finland, covering a total of 483,000 km (Finnish Transport Agency, 2013)

The extensive Social Statistics Data Warehouse of Statistics Finland offers dwelling coordinates and work place coordinates for citizens along with a variety of demographic features (Nieminen, J. & Myrskylä, P., 2010). Work place coordinate coverage was 91.2% of all inhabitants in 2010 and 92.4% in 2008. In future, it is expected, that fewer people will have a fixed singular coordinate pair due to the changing nature of the "work place".

The Grid Database is based on this warehouse. The minimum size of a grid is 250 square meters, and for each grid the data include the population structure, educational structure, consumer structure, and many kinds of building and housing information, besides workplace and service structures (Statistics Finland, 2012). The coverage of map coordinates is 99% of all inhabitants.

2. Pairwise computing of commuting distances

Modelling accessibility is among the fastest growing areas in GIS research. Often, a high quality route network is needed in order to calculate both travel distances and travel time between specific points. In the first case, a straightforward approach for using Digiroad in the official statistics is given: pairwise commuting distances for the approximately 2.1 million employed Finns that have coordinate pairs. The data required for such calculations are two years behind the actual target year.

The actual data for computations can be reduced by simple editing using linear distances. However, in this case, all the distances with a linear distance of more than 0.3 kilometres were included (60,000 less than the total).

The ESRI ArcGIS software is used, of which Network Analyst Route Solver is called with a Python code. The method used is a common Dijkstra's algorithm with hierarchical routing (ESRI, 2005), which makes calculation faster and is seen as a natural option here. The impedance attribute is the length of a route and the hierarchy is a seven-class functional classification attribute (Class I main road, Class II main road, Regional road etc.).

Table 1. *Distance statistics in meters: median, mean of 0-200 km distances, 25th percentile, 75th percentile and quartile coefficient of dispersion (Q3 - Q1) / (Q3 + Q1).*

Type	Median	Mean	Q1	Q3	QCD
Linear	6,019	13,342	2,047	15,216	0.763
Route	7,655	15,918	2,670	18,644	0.749

3. Commuting distance results

The median ratio of a route distance and corresponding point-to-point linear distance between a home and a main workplace is 1.22, the mean of ratios is 1.28, and the standard deviation is 0.64 meters (after the imputation procedure of the shortest routes, under 300 meters). Naturally, the ratio changes depending on the length class and geographical area. Especially Finnish archipelago areas and the Finnish Lakeland covering a quarter of the total area tend to have higher ratios.

As Table 1 indicates, the distribution of the commuting distance is skewed, skewness being 7.42. Thus, the basic descriptive statistics shown is based on very robust measures: quartiles. Specifically, the dispersion measure is given as a quartile coefficient of dispersion (Bonett, D. G., 2006): the ratio of the interquartile range and the interquartile sum:

$$QCD = \frac{Q_3 - Q_1}{Q_3 + Q_1}.$$

This is one robust and simple measure of variation that is used to make comparisons between sets of data.

Figure 1 shows the area differences in commuting distances. Many remote areas tend to have shorter distances on average but the deviation is very high. The metropolitan area of Helsinki, Helsinki Sub-region at LAU 1 level, indicated by a black square, which covers 29.7% of the total employed population of Finland, has longer commuting distances than the average (not to mention the commuting time). The median for this metropolitan region is 9.83 km.

Commuting distances for the employed population are now part of the Social Statistics Data Warehouse for the years 2008 and 2010, and soon for 2011. An obvious continuum for the study itself is estimating commuting times. Speed limits and different kinds of functional classifications for the road elements are not enough when dealing with commuting related to rush hour traffic.

A research group on Spatial Patterns on Accessibility of the University of Helsinki has done extensive research and produced free software that calculates the travel time from any selected point to another within the Helsinki Metropolitan Area. It takes into account the time of day, traffic lights and intersections, even the time that parking takes (Salonen M. & Toivonen T., 2013 and MetropAccess, 2013).

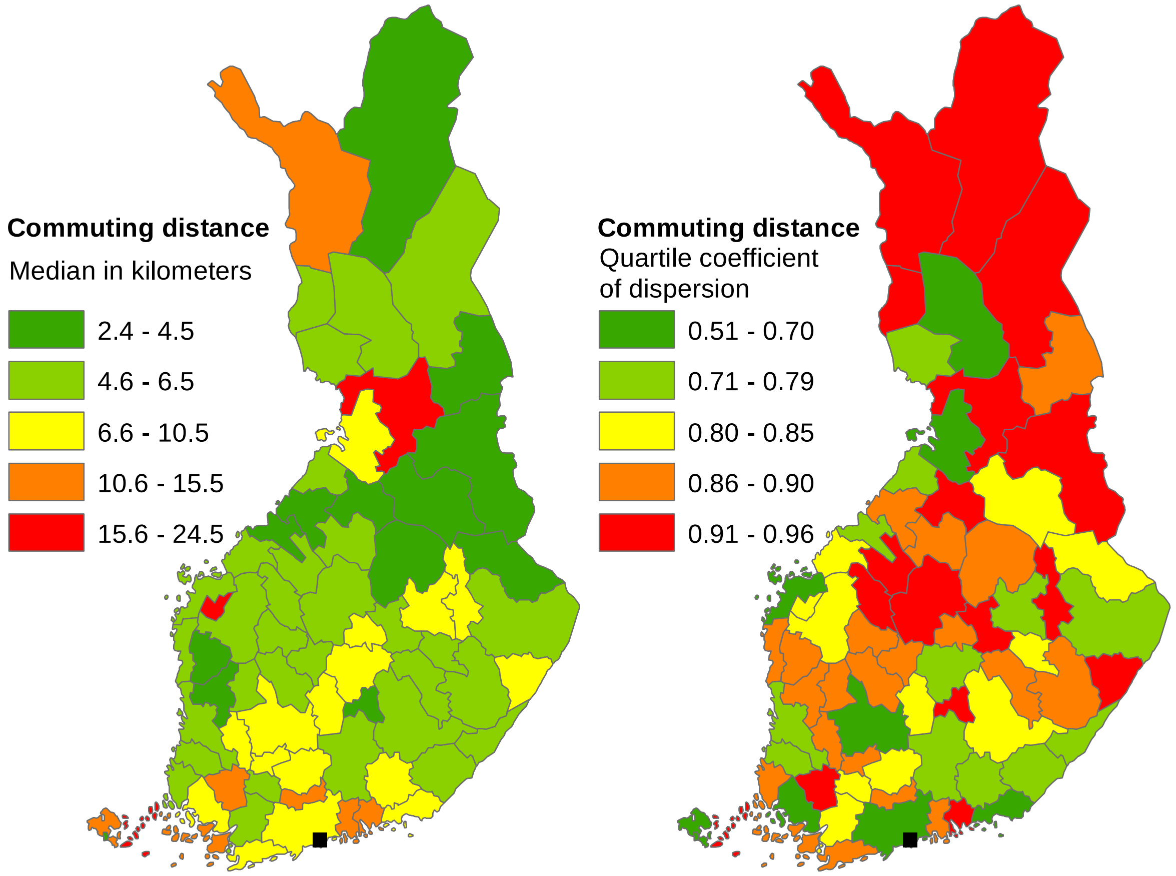

Figure 1. *The median and the quartile coefficient of dispersion of commuting distances along the route network for populations of the sub-regions of Finland, at LAU 1 level, in 2010. Helsinki is marked with a black square marking also the surrounding Helsinki Sub-region. Classification is based on the Jenks optimization method.*

4. Measuring travel time: the case of public hospitals

Commuting time estimation clearly comes with many complexities but in some other applications, estimation can be more straightforward. A case study on accessibility to public 24/7 hospitals of Mainland Finland is one example. This is a part of a study for an ongoing ESSnet Geostat 1B -project (Eurostat Grant Agreement no 50502.2009.004-2011.536).

The data, besides the Grid Database and Digiroad, is based on a study by Reissell, E. et al. (2012), where 24/7 health centres' services in Mainland Finland were identified. Locations of certain emergency rooms were geocoded based on their address information.

The idea was to use a simple approach to travel time that is the estimation of travel time using speed limits for each traffic element (smallest unit of Digiroad's centre line geometry). When speed limit information is not available, it is estimated using the functionality class of a traffic element.

The 30 minutes and 60 minutes detailed service areas for each unit (ESRI, 2005) were created by using a similar route optimisation algorithm as in the previously presented method but without hierarchical routing.

In the service area layer, a grid is selected as part of the area if it intersects the service area polygon. In this case, the selection procedure itself is not expected to increase bias due to the detailed polygons and the size of a grid. Hospital districts are only shown for informative purposes and are not part of the selection procedure itself.

5. Results for the emergency service area coverage

As shown in Table 2 and Figure 2, travel times to the nearest hospitals vary by area; travel times are higher in sparsely inhabited areas. However, 88.5% of the entire population is within less than 30 minutes coverage and almost all, 97.3%, should get to the nearest emergency facility under 60 minutes. Results also show some typical variation between different age groups: pensioners are living further away from the hospitals and health centres than the younger generation.

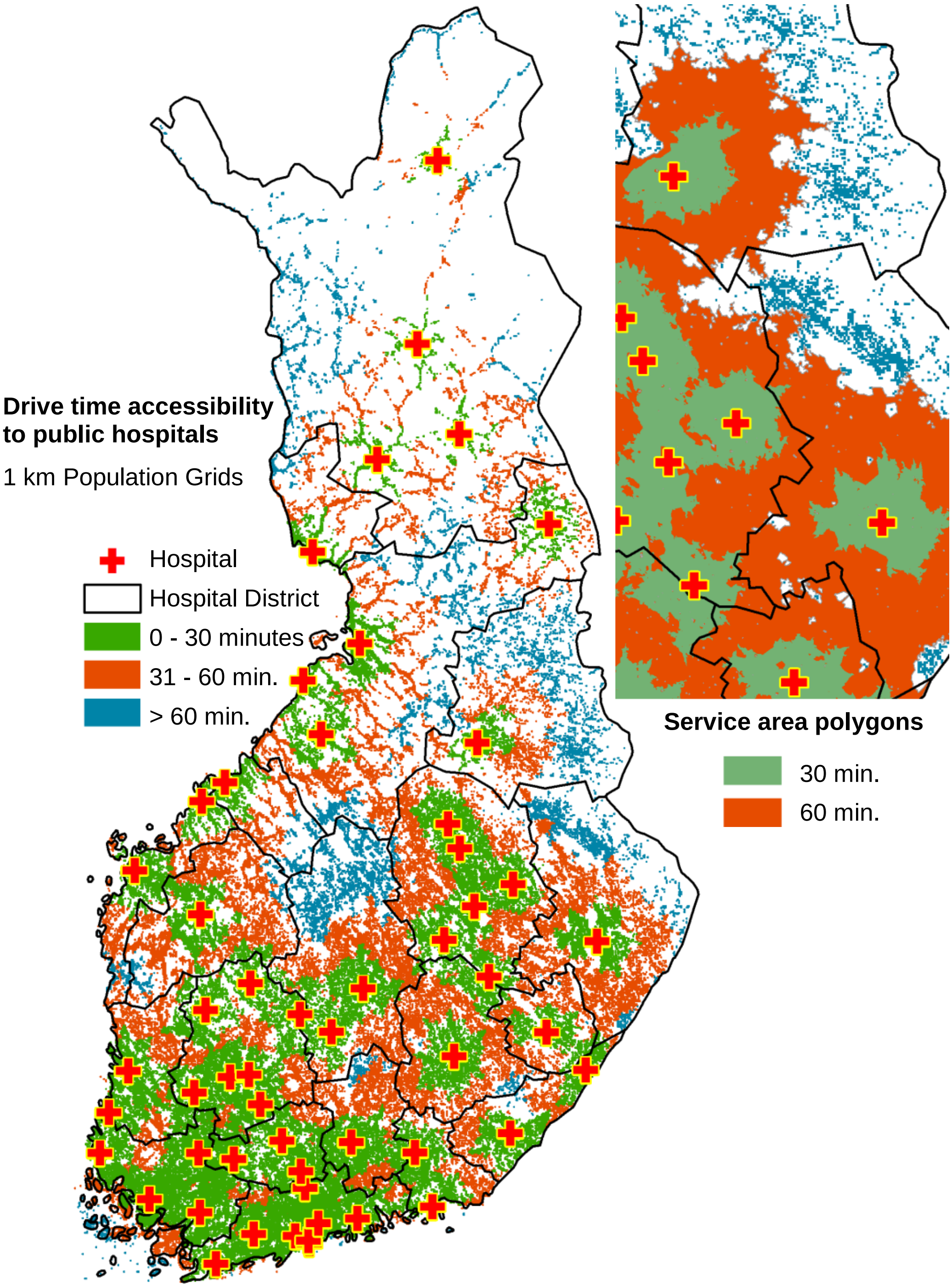

Figure 2. *Drive time from 1 km x 1 km (2012) permanent population grids to the public hospitals, measured with an example of the corresponding service area polygons.*

Table 2. *Relative 30 minutes drive time coverage of public hospitals and health centres with 24/7 emergency rooms, by age groups.*

Hospital district	1-15	16-64	65+	All
All	0.896	0.894	0.846	0.885
Min, Kainuu	0.591	0.581	0.519	0.570
Max, Helsinki	0.991	0.991	0.986	0.991

6. When distance could be enough: distances to elementary schools

A relevant part of the accessibility studies are the service areas of elementary schools at municipality level. Obviously, travel time can be important when cars are being used, otherwise distance is enough.

Another approach for the service areas is to estimate the distances directly from each dwelling unit, which, in this case, are aggregated to 250 m x 250 m cells. Schools with relevant attributes are picked from the Register of the Educational Institutions by Statistics Finland.

Finland has a compulsory comprehensive school system in which buildings can be classified into three distinctive categories: lower comprehensive schools (children aged 7 to 12), upper comprehensive schools (children aged 13 to 15) and general schools including both. Comprehensive school level special education schools, 3.5% of all the schools, are excluded.

Respectively distances are measured, and weighted accordingly, separately for those cells that are populated by children aged 7 to 12 and by teenagers aged 13 to 15. Obviously, the nearest school must be located in the same municipality.

Similar to the previously described algorithm, ESRI ArcGIS Route Analysis Tool is implemented in an efficient way by creating an origin-destination cost matrix layer for Digiroad but non-hierarchically also allowing pedestrian routes to be part of the route optimisation process.

Table 3 shows the results for the five largest municipalities. More generally, most schoolchildren are living within 3 kilometres from their nearest schools. Naturally, these figures vary widely by municipality, in other words, by the area and by the density of the population.

Table 3. *Route distances to the nearest lower comprehensive school (for the permanent population aged 7 to 12 in 2012) and to the nearest upper comprehensive school (for those aged 13 to 15) in the five largest municipalities of Finland: relative frequencies and total populations in each class.*

Municipality	<1 km	< 3 km	< 5km	Population
Helsinki lower s.	0.811	0.996	1.000	29,246
Helsinki upper s.	0.552	0.988	1.000	14,647
Espoo lower s.	0.669	0.984	0.994	18,799
Espoo upper s.	0.417	0.934	0.982	8,779
Tampere lower s.	0.583	0.967	0.984	10,558
Tampere upper s.	0.368	0.913	0.977	5,446
Vantaa lower s.	0.607	0.983	0.994	13,928
Vantaa upper s.	0.348	0.887	0.974	6,780
Oulu lower s.	0.459	0.966	0.982	8,969
Oulu upper s.	0.329	0.848	0.931	4,206

7. Further research and conclusions

As mentioned earlier, travel time is a highly relevant part of accessibility studies and further research is needed. Many accessibility challenges are related to rush hour traffic. Hence, for example, the estimation of the commuting travel time becomes much more complex than the general 24/7 emergency accessibility.

The extensive and detailed national road network has been proven useful in population statistics and will be seen as part of the social statistics data warehouse based applications in forthcoming years.

References

Bonett, D. G. (2006). Confidence interval for a coefficient of quartile variation. Computational Statistics & Data Analysis 50 (11), 2953–2957.

ESRI (2005). White Papers for ArcGIS Network Analyst. *http://support.esri.com/en/knowledgebase/whitepapers/list/productid/87*

Finnish Transport Agency (2013). Digiroad, Description of Data Objects v. 3.4. *http://www.digiroad.fi/dokumentit/en_GB/documents/*

MetropAccess (2013). Saavutettavuuslaskentaa autoilijan näkökulmasta: MetropAccess-Digiroad. University of Helsinki. *http://blogs.helsinki.fi/saavutettavuus/metropaccess-digiroad/*

Nieminen, J. & Myrskylä, P. (2010). Census data warehouse - use of census data. Working paper 18 for the Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses, The Hague. *http://www.unece.org/stats/documents/2010.05.census.html*

Reissell, E., Kokko S., Milen A., Pekurinen, M., Pitkänen, N., Blomgren S. & Erhola, M. (2012). Sosiaali- ja terveydenhuollon päivystys Suomessa 2011. National Institute for Health and Welfare (THL). Report 30/2012.

Salonen, M. & Toivonen, T. (2013). Modelling travel time in urban networks: comparable measures for private car and public transport. Journal of Transport Geography 31, 143–153.

Statistics Finland (2012). Grid Database 2012. *http://www.stat.fi/tup/ruututietokanta/griddb2012_description.pdf*