MORAVIAN GEOGRAPHICAL REPORTS

sciendo

The Czech Academy of Sciences, Institute of Geonics
Palacký University Olomouc, Faculty of Science

journal homepage: www.geonika.cz/mgr.html

doi: https://doi.org/10.2478/mgr-2024-0021

How data validity and granularity affect jobs accessibility: A case study from the Czech Republic

Daniel FRANKE [a*], Tomáš PELTAN [a], Vojtěch NOVOTNÝ [a]

Abstract

Potential accessibility is a commonly used tool for estimating the impact of changes in land-use and transport infrastructure. The accessibility is derived from a spatial interaction model and two components are crucial for its calculation: land-use and transport. While data relating to the transport component are often well available with a good validity, data representing the land-use component are more challenging. Using a selected case study area from the Czech Republic, the aim of this paper is to explore the sensitivity of the job-related accessibility calculation to different data on the land-use component, in order to explore and possibly overcome these problems with validity. Two different datasets were compared: valid jobs data based on available population and commuting data from the Census of the Czech Statistical Office with very coarse aggregation, and land-use data from the Urban Atlas with low validity of what is measured but very fine spatial aggregation. The analysis of the maps reveals clear spatial patterns and suggests some limitations when using alternative data. The comparisons based on the Urban Atlas data at distinct levels of aggregation reveal problems in some areas of the hinterland of Prague and support the need for more detailed data in these areas and specific types of analysis that may be sensitive to this type of error.

Keywords: *Potential accessibility, jobs, alternative data, Urban Atlas, Czech Republic*

Article history: *Received 1 December 2023, Accepted 6 November 2024, Published 31 December 2024*

1. Introduction

Accessibility and transport costs play an important role in spatial decision making, as confirmed by a number of models over time using different methodologies, ranging from monocentric models of regions through spatial interaction models to more advanced simulations (von Thünen, 1826; Alonso, 1964; Lowry, 1964; Waddell, 2002; Lopane et al., 2023). Despite some serious criticism, this approach has proved to be a useful tool for modelling both urban and regional development (de la Barra et al., 1984; Echenique, 1994; Modelistica, 2011) and for analysing spatial attractiveness. It is often used as a tool for evaluating policies related to transport infrastructure (Vickerman et al., 1999; Maier et al., 2007; Spiekermann & Wegener, 2007), with accessibility to jobs being considered an important factor in residential location choice (Ahlfeldt, 2007). Accessibility is also crucial for equity (Wee, 2022) and spatial justice (Nyamai & Schramm, 2023).

The estimation of jobs accessibility is limited by the validity of available data, in particular when using only publicly available datasets. Validity of data is related to the use of data and reflects whether the data correctly represent relevant aspects of reality. The validity in spatial analysis is usually concerned with (1) what is measured and (2) how is it aggregated and (mis)allocated. From the first point of view data are valid if they provide direct information about jobs and their types. Data providing e.g. areas of jobs-related land uses are only an indirect indicator with much lower validity as the number of jobs per square hectare can differ.

But validity in spatial analysis is not limited to what is measured, but also depends on the level of aggregation and where the data point is located in space. Coarse aggregation and substantial difference between the real location of jobs and the location of the data point can severely affect results of analysis. This is important particularly in the analysis of infrastructure improvement where the real location of jobs and the location of the aggregated data point can result in very low validity of such data and results.

This problem becomes more important as core municipalities of urban regions become more polycentric and their administrative boundaries extend during their growth while the data are still represented by a single point located in the city centre. This problem is exacerbated by growing urban congestion or in cases when e.g. public transit is the mode of choice. The difference between real time distance and the distance to the data point can easily differ by more than half an hour moving the jobs beyond practical accessibility. The misallocation is also affecting the evaluation of transportation network improvements which often include building ring roads which connect housing areas to jobs in secondary urban centres. When the analysis misrepresents the

[a] Department of Landscape and Urban Planning, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic (*corresponding author: D. Franke, e-mail: *franke@fzp.czu.cz*)

location of jobs, it renders these improvements useless while they are important in real-life situations. See chapter 3 for an example from Prague area.

The existing publicly available datasets are either valid from the first point of view (e.g. the Czech Statistical Office census data) but have very low validity from the second point of view (coarse aggregation) or have high validity from the second point of view (individual areas of land uses) but have very low validity concerning of what is measured (i.e. jobs-related land use areas rather than direct number of jobs).

The main objective of this paper is to compare datasets of both types and examine the possibility of using Urban Atlas (Logan et al., 2017) land-use data, which have finer spatial granularity but are less valid, instead of Census job count data provided by the Czech Statistical Office, which provide valid numbers of jobs but have very coarse granularity in their publicly available form. The paper focuses on two research questions: how is potential accessibility of jobs affected by datasets with different validity and how is it affected by different levels of data aggregation?

2. Theoretical background

2.1 The basic concept of potential accessibility

The accessibility is considered an important measure of attractiveness with wide application in both spatial analysis and land-use and transport modelling (LUTI) (Alonso, 1964; Lowry, 1964; Vickerman et al., 1999; Maier et al., 2010; Modelistica, 2012; Echenique, 2011). It is also a valuable tool for policy evaluation and decision support, such as the impact of energy price increases and possible policy measures to address this issue (Spiekermann & Wegener, 2007) or potential accessibility of secondary schools (Rosik et al., 2021). The attractiveness of the area in a broader sense is also determined by other various factors. Many of these factors are non-spatial in their nature (e.g. the intra-family ties described by Musil, 1966), while most other factors influence location choice only on a small scale (e.g. socio-demographic characteristics of the neighbourhood, access to local green spaces, playgrounds, amenities, etc.). These factors are therefore of limited use in measuring attractiveness at the regional level. As a result, attractiveness in a narrower sense is estimated as a function of accessibility to a variety of amenities and other destinations, depending on the purpose of the analysis, while other factors are considered to be a random or residual component of overall attractiveness. This approach is useful for the evaluation of transport policies, while it is of limited use in some other cases, such as the analysis of social segregation, the impact of clean air policies or the institutional capacity of exporting firms in different regions.

Geurs and van Wee (2004) describe four types of accessibility measures: infrastructure-based, location-based, person-based and utility-based. This paper deals with potential accessibility as an example of a location-based accessibility measure. Different measures of potential accessibility affect the attractiveness for different land-uses. Population potential is recognised as an important location factor for commercial firms as it is a measure of potential market size (Ahlfeldt, 2007). The accessibility to jobs has been proposed as a measure of residential attractiveness (Alonso, 1964) and has been used as a location factor in land-use transport interaction models. On a smaller scale, accessibility to amenities and green spaces plays a significant role. Some less quantifiable accessibility measures are also important, such as the accessibility of CBDs (central business districts) and other city centres (Hansen, 2009). Since its early application, the measurement of accessibility has been based on two basic principles: cost/disutility and gravity models. The first approach is based on economics with the estimation of transport costs and economic disutility (von Thünen, 1826; Alonso, 1964; Echenique, 1994; Modelistica, 2012), while the other approach is based on a more descriptive attitude based on analogy with gravity theory and entropy (Lowry, 1964; Vickerman et al., 1999) or a more intuitive interpretation of (El-Geneidy & Levinson, 2006), who suggest that measuring accessibility requires knowledge of the degree of attraction and repulsion between destinations.

Our research follows the latter approach, and our measure of accessibility is derived from the concept of potential accessibility (Spiekermann & Wegener, 2007), also known as integral accessibility (Morris et al., 1979). It is measured as the sum of discounted weights of potential destinations with the general form of Equation 1:

$$A_O = \sum_{d \in D} (W_d * DDF(o, d))$$

where o = location of origin, d = location of destination, D = set of all destinations, W_d = weight of the destination d and $DDF(o, d)$ = distance decay factor between origin o and destination d.

The weight measure is a numerical expression of the attractiveness of the destinations that can be reached from the location. Its value largely depends on the type of accessibility being measured (i.e. what is to be reached), and as such it can be the number of jobs, retail space, population or whatever else the research needs. The distance decay function reflects the spatial 'friction' between the two locations, which reduces the relative importance of destinations further away from the location being studied.

2.2 The transport component of potential accessibility

The transport component affects the ability to reach specific destinations at a specific cost (e.g. time or money) and thus reflects the mobility available in different locations. In assessing potential accessibility, the transport component affects the distance decay factor – a function that represents spatial friction, leading to a decrease in the importance of destinations that are not easily accessible.

There are two key issues in estimating the value of the distance decay factor: what measure of distance to use and what functional form to use. Distance is often measured either in spatial terms (e.g. distance in kilometres) or in terms of time. The former is appropriate in cases where either monetary transport costs are a major part of transport disutility, or for modes of transport with small differences between average and actual speed (e.g. walking or driving a car in an uncongested environment with a substantially and equally limited maximum speed). The distance measure has been applied in both the classical works (von Thünen, 1826; Alonso, 1964; Lowry, 1964) and more recent works such as Ahlfeldt (2007).

Travel time has become the key component of travel disutility as travel costs decrease and speed differences (e.g. between high-speed and conventional rail, car traffic on motorways and other roads, etc.) and congestion increase. Time distance has recently become the most common measure for estimating accessibility (e.g. Vickerman et al., 1999; Maier et al., 2007). Monetary expression of costs and disutility is another alternative to measure distance, often based on both spatial and temporal distance-related costs. This approach is used, for example, in the Tranus model (Modelistica, 2012). Spatial friction can be described by several functional forms. The early use of accessibility by von Thünen (1826) and Alonso (1964) measured distance decay or friction as a simple cost function based on the assumption that transport costs are linear in distance travelled and specific to different land-use. The land-use component is represented just by one point (centre, central business district), so accessibility was equal to the (dis)utility of transport, without the need to compare the influence of more destinations with different parameters.

Later work was inspired by physics, namely Isaac Newton's gravitational formula, where each target's 'weight' is discounted with a distance decay parameter in the form of an inverse power function. This form is still widely used, with Vickerman et al. (1999) being an example. Another frequently used functional form is the exponential function in the form $exp(-a.d_{o,d})$, where the parameter a reflects the mobility (and hence the steepness of the decay function) and $d_{o,d}$ is the distance between the origin and destination locations. This form of distance decay function is used, for example, by Ahlfeldt (2007). There are also ongoing efforts to find functional forms with a better fit to real data, e.g. de Vries et al. (2009) and Novotný (2011).

2.3 The land-use component of potential accessibility

The main difference between mobility and potential accessibility is the relationship with the land-use component. Accessibility is influenced by the spatial arrangement of potential destinations (e.g. amenities, jobs, potential labour force), and thus the places with low mobility close to important destinations (e.g. congested city centres) may have very good accessibility while mobility may remain low.

Increased mobility, much of it provided by the private car, affects the spatial location of some types of destinations (e.g. retail and amenities) and increases the dependence of accessibility on mobility (Hanson & Giuliano, 2004). A comparative analysis of different urban areas reveals a significant impact of land use configurations on accessibility.

For example, review by Chatman and Nolan (2011) demonstrates that urban centres and areas near public transport stations typically exhibit higher density and accessibility scores with multiple effects on agglomeration economies on local, city-wide and national scales. The relationship between land-use and transport systems has been the subject of considerable research, with Boarnet and Crane (2001), for example, demonstrating that integrating land-use planning with transport infrastructure can improve accessibility. Their work is an example of how accessible transport networks can extend the reach of different land-uses, and thus enhance the relationship.

The land-use component data for calculation of jobs accessibility available in the Czech Republic often have serious shortcomings which limit their use. We expect the land-use component data to allow for international comparisons, to be available at least for the main functional urban areas (FUAs), and the aggregation shouldn't be a source of significant error. The data should be the most direct indicator of the importance of the job (preferably the number of employees actually working in the destination) and should be updated at reasonable intervals with minimal time lags (time between when the data are collected and when the data are available to researchers). The data should also be open accessed (available to all researchers [free as in "freedom"] and at reasonable cost [free as in "free beer" being the optimum]).

2.4 Aggregation and the modifiable areal unit problem

The size and granularity of the zones has been considered another important factor in accessibility modelling, with results being affected by both the size of the zones and the division of the area into zones (Kwan & Weber, 2008). This problem is known as the modifiable areal unit problem (MAUP). There are two strategies for dealing with the modifiable areal unit problem or the scale of analysis. The first strategy addresses the zoning/aggregation itself – either by using the disaggregated data (e.g. individual dwellings/workplaces) or by using the correct scale and division into zones, the approach advocated by e.g. Openshaw (1983) who suggests solving the automatic zoning problem as part of the analysis. The other solution to this problem is based on the choice of scale-independent accessibility measures, e.g. the space-time accessibility measures adopted by Kwan and Weber (2008).

While the estimation of the destination weight parameter seems to be rather resistant to aggregation effects, the distance parameter is affected by zoning in many ways. The error is caused by replacing the set of basic units with an uneven distribution within a single point representing the whole zone. The centre of the zone has been often used as a proxy for the zone. The problem of distance measurement error has been investigated in location assignment research and three sources of error have been identified in measuring the distance between the zone and the point outside the zone (including other zones; type A error), the self-distance error (type B) and the misallocation error (type C). Hewko et al. (2002) analysed the impact of these origin-side errors on the measurement of amenity accessibility and concluded that the impact of the zoning-related error decreases with the service area and increases with the density of the amenity type. They also suggest some strategies to overcome the sources of error when more disaggregated data are available: either to use the better estimate of the zone reference point (replacing the centroid with the weighted average of the basic units) or to calculate the average distance as the weighted average between the basic units. The first method is more data intensive, while the second method also requires much more computing power (*ibid.*). While Hewko et al. (2002) focused on origins, the symmetric nature of the origin-destination relationship allows the results to be extended to the destination side as well.

3. Data and methods

3.1 Study area

The choice of study area reflects our research questions and aims. The Prague capital is the largest municipality in the Czech Republic with 1,301,432 inhabitants according to 2021 Census (CZSO, 2021), i.e. more than 12 percent of the total population of the Czech Republic. The data on jobs are available as a single data point located in the centre of the municipality.

This coarse granularity causes substantial problems in accessibility analysis. Consider an evaluation of planned northern ring road around Prague (the D0 highway). We would expect improvements of jobs accessibility in the northern hinterland of Prague (e.g. in Klecany town) because the accessibility of jobs in Dejvice city quarter and at the Václav Havel Airport will substantially improve (see Fig. 1). If we use a dataset that represents all jobs in Prague with one data point located in the centre of the city as a basis for our analysis, the results will be quite far from valid as the ring road will not improve the ability to reach the city centre. The importance of validity related to these issues will grow as we switch to slower modes of transportation, e.g. public transit, cycling and walking.

This problem was a major factor for the selection of the study area which includes Prague and its hinterland. We compared both datasets and decided to delineate the study area to the maximal continuous extent of available datasets. The study area therefore consists of three contiguous regions of the Czech Republic – the metropolitan region of Prague and two regions of level 1 meso-regional centres – the regions of Pilsen and Hradec Králové-Pardubice in terms of functional interactions (see Erlebach et al., 2016). Administratively, the study area is part of five NUTS 3 regions, including Capital City of Prague, Central Bohemia, Hradec Králové, Pardubice and Pilsen (see Fig. 2).

The extent of the study area in all regions was limited to areas where both datasets were available, with the Urban Atlas being the limiting factor. This results in the boundaries following functional urban regions as derived from commuter flows rather than administrative boundaries of districts (see Maier et al., 2007 for more details on functional urban regions in the Czech Republic).

Municipalities outside these borders were not included in the calculations in order to avoid the error associated with different extents of the study areas.

3.2 Data and its regional specifics

The road system is described using the Open Street Map Volunteered Geographic Information (VGI) dataset (OSM, 2011). The Open Street Map data is publicly available under the Creative Commons licence, and comparisons with commercial datasets in the UK (Haklay, 2010) and Germany (Neis et al., 2011) suggest that it is fairly reliable and up-to-date, with some spatial variation in completeness and accuracy. Two data sources were used to calculate the weight measure. Data on the population and its behaviour were obtained from the Czech Statistical Office, which provides information on the economically active population (employed, unemployed) at the municipal level. Data from the Czech Statistical Office on commuter flows between municipalities were also used. We use data from 2011 (CZSO, 2011) which are closest to the Urban Atlas dataset from 2012. Data on jobs-related land-uses for a more detailed distribution of jobs within larger employment centres were obtained from the Urban Atlas, distributed online by the European Environment Agency. We use Urban Atlas data from 2012 (Urban Atlas, 2012). The Urban Atlas provides European land-use and land cover data for large urban areas. In case of future use of the methodology, it will be possible to combine newer datasets, namely the 2021 Census of Population, Houses and Dwellings of the Czech Statistical Office (CZSO, 2021) and the latest Urban Atlas, which is being created for the period 2021 to 2024 within the Copernicus Land Monitoring Service.

Accessibility modelling in the Czech Republic is often based on census data provided by the Czech Statistical Office (CZSO), as these have been the only publicly available data on jobs and population. These data are available for all municipalities in the

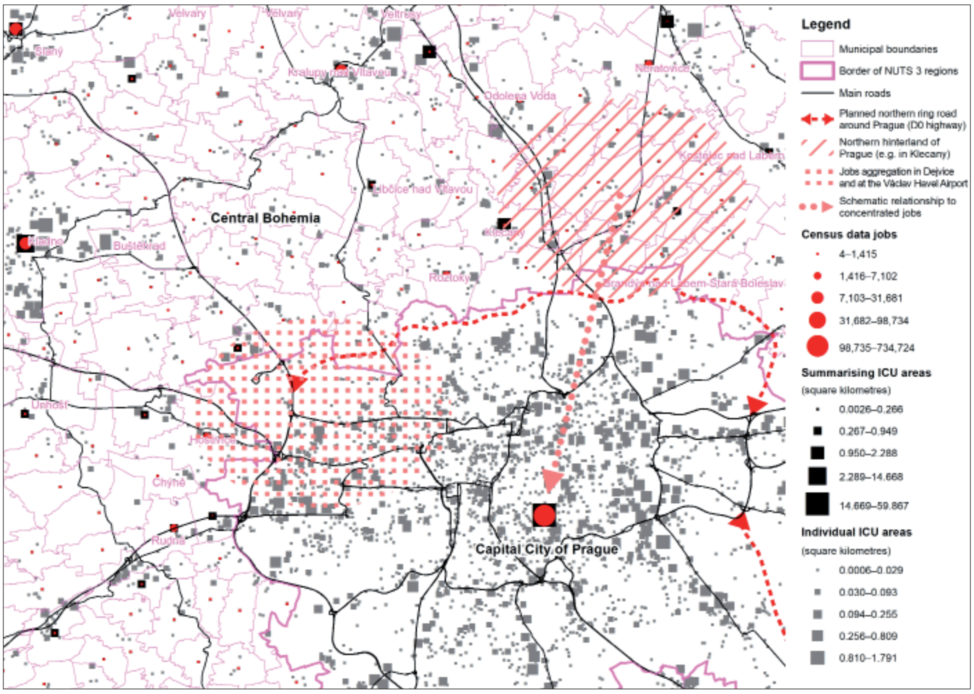


Fig. 1: Coarse aggregation and substantial difference between the real location of jobs and the location of the data point
Source: Authors' elaboration

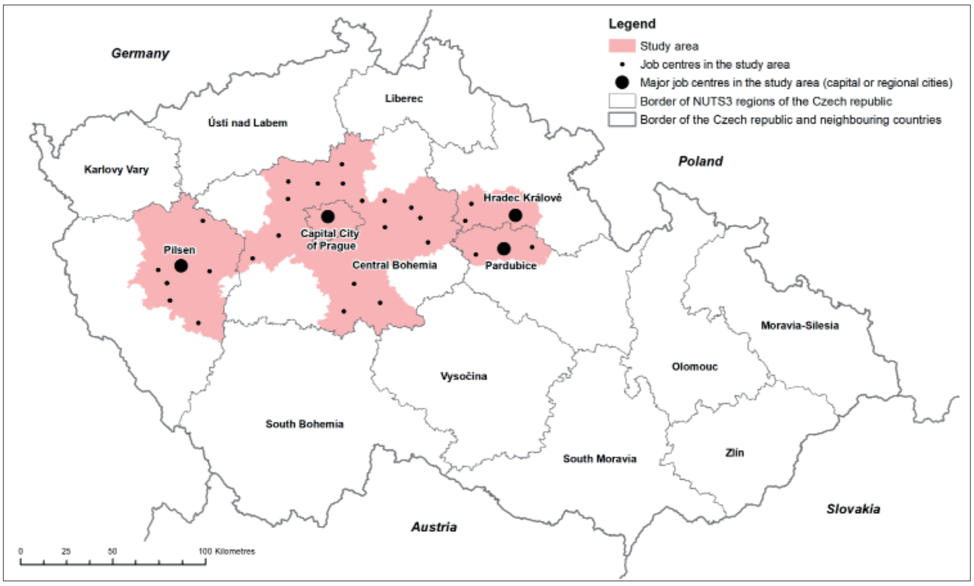


Fig. 2: The study area within the Czech Republic
Source: Authors' elaboration

Czech Republic as part of two datasets. The first dataset contains data on the population, including data on the economically active and unemployed population. The other dataset describes all commuter flows between municipalities. The combination of both datasets makes it possible to calculate the number of Census data jobs in each municipality. The advantage of these data is their reliability and validity, as jobs are directly measured and the whole population is surveyed. The CZSO census data have some shortcomings. The main obstacles are the limited availability (only in the Czech Republic), a long period between censuses and the level of their aggregation. The lack of comparable international data prevents direct international comparisons and leads to errors in the evaluation of regions close to borders. The 10-year census period is often extended by long processing times – data on commuter flows from the 2011 census were not available until the end of 2013, data from the 2021 census were not available until mid-2023.

The level of aggregation is high, with a very coarse spatial granularity, as data are generally available only at the level of municipalities, with the exception of public administration. The capital city of Prague (NUTS 2 region, more than 1.3 million inhabitants, 496 square kilometres and over 25% of the GDP of the Czech Republic (IPR, 2020) is the notorious example as it is represented with one data point as a single municipality with missing data of its LAU level 1 and 2 sub-municipalities.

Similar data are provided by the Ministry of Finance since at least 2000 (last data MoF, 2024) as part of the calculation of the percentage share of individual municipalities in the shares of the national gross value added tax and income tax revenue. This calculation includes the value of the number of employees. This dataset is updated every year and thus overcomes the obsolescence of the census data, but the problem with data aggregation at municipality level is the same. The alternative data describe either the land-use or the census buildings including their use. CZSO data on buildings have a much finer granularity, but international availability is limited, and the data are also closed-access. The data are incomplete; the floor areas needed to estimate the size of buildings and the number of jobs are the only available for buildings completed after 2001, which makes the data rather useless.

The Urban Atlas land-use data are based on remote sensing and have the advantage of free access and international availability for large urban areas in the EU. The data describe land-use at a fine territorial granularity, overcoming the disadvantage of CZSO data in larger municipalities. The main weakness of the data is their relatively weak relationship with jobs per se, as these data don't take into account the density of land-use (such as built-up areas, floor space and jobs per square kilometre). This makes it a limited proxy for the number of jobs, as areas of a similar type may have very different numbers of jobs, for example due to temporary under-utilisation, computerisation of manufacturing resulting in low numbers of jobs per floor area, or different sizes of office buildings within the same built-up area. The data are limited to the main city regions (functional urban regions) and therefore they do not cover the whole territory of the Czech Republic.

It might also be the case that the defined area of each city could change over update cycles, which could raise questions about the comparability of analyses over longer time periods. This issue is partially resolved through a comparison of Urban Atlas data with Census data at the municipal level. This approach enables the flexible addition and removal of municipalities according to the defined area based on Urban Atlas data.

The Urban Atlas is not the only source of land-use data (Pazúr et al., 2017; Micek et al., 2020). The ZABAGED dataset is provided by the Czech Office for Surveying, Mapping and Cadastre with limited availability for the area of the Czech Republic. Landsat data are another source of land-use data available (not only) for the Czech Republic. See Table 1 for further details.

3.3 Methods

The jobs accessibility is an indicator of the attractiveness of an area that is expressed by the supply of potential commuting destinations (i.e. jobs areas). It is determined by the ability to reach the jobs areas using the transport infrastructure. The result of the calculation is a dimensionless value that provides information about the attractiveness of the area. The higher the calculated value, the more attractive an area is. The results are normalised to the maximum value on a scale of 0 to 100 (where 100 is the maximum value). The potential jobs accessibility indicator is calculated for municipalities using the formula of Maier et al. (2010):

$$P_o = \sum_{i=1}^{n} (W_i * DDF(t_{oi}))$$

where P_O = potential accessibility in the municipality *O,* W_i = weight of an employment area *i* (see further), t_{Oi} = time distance between origin *O* and destination *i* and DDF() = distance decay function based on Novotný (2011)

We use the distance decay function based on the cumulative normal distribution (CND, Φ). Recent research by Novotný (2011) suggests that the function based on cumulative normal distribution is a more appropriate form for describing people's willingness to travel for a given time distance, and therefore a more relevant form for the distance decay function. Based on the tested values and compared to similar studies worldwide the final values of mean (μ) and standard deviation (σ) are determined: $\mu = 35$ minutes and $\sigma = 20$ minutes. The normal distribution function is converted to a function where with increasing time in minutes the

Characteristics	CZSO Census data	Urban Atlas data	ZABAGED data	Landsat data	Ministry of Finance data
Availability in the Czech Republic	whole	large city regions	whole	whole	whole
International availability	no	yes	no	yes	no
Indicator type	jobs	land-use	land-use	built-up area	number of employees
Granularity	municipality	polygons	area with the same land-use	area with the same land-use	municipality
Validity regarding aggregation and misallocation	good (small municipalities) to very low (large municipalities)	high	high	high	good (small municipalities) to very low (large municipalities)
Update period	10 years (last 2021)	every 6 years and partly continuously	continuously	continuously	continuously since 2000
Paid data	no	no	no (from 7/2023)	no	no

Tab. 1: Data comparison characteristic
Source: Authors' compilation

willingness to commute to the centres decreases from 100 percent at 0 minutes to about 5 percent at 72 minutes, using the following equation (Novotný, 2011; Franke et al., 2017):

$$W(t_{od_i}) = (1 - \Phi[(t - \mu)/\sigma])$$

Time-distances between individual centres were calculated from the physical distances between them, assuming that drivers drive within the speed limit and are further limited by factors of the geometry of the given road, such as slopes and curvilinearity. The calculation was carried out using ArcGIS with the Network Analyst extension to create a routable network dataset based on road network from Open Street Map.

Residential roads and streets are defined by built-up areas where speed limits are different in the city (speedways – 70 km/h, main roads – 45 km/h, other roads – 40 km/h) and in rural areas (main roads – 50 km/h, other roads – 45 km/h). The Czech technical standard CSN 73 6101 determines the influence of curvature on the design speed of roads in relation to the road category:

$$K = \frac{\sum_{i=-1}^{j} |\gamma_i|}{l}$$

where K = curvilinearity in grad per km, γ_i = the angular change in grad, l = the length of road in km and j = the number of sub-sections of the road.

Categories of roads and its speed limits are derived from design speed as shown in Table 2. The lowest of the speed limits determined by legislation, slope and curvilinearity was used in the calculation (see Tab. 3).

The location of municipalities is approximated using the definition points of the Czech Statistical Office. The calculation of the weight factor of an employment centre is one of the most important factors influencing the attraction and resistance between destinations (Spiekermann & Wegener, 2007). Three options are used to calculate the weight of a job area, which are later compared in a case study. The first option is to calculate the number of Census data jobs in the destination area using data from the Czech Statistical Office's census:

$$F_d = EA - EAU + I - O$$

where F = weight of a destination area, d = destination area, EA = economically active population in the destination area, EAU = economically active but unemployed population in the destination area, I = daily commuter incoming to destination area and O = daily commuter outgoing from destination area.

The second option for calculating the weight of a job area is based on the Urban Atlas land-use data. In this option, we summarise the area of all ICU (industrial, commercial, public and private units) derived from Urban Atlas at the municipality level:

$$F_d = \sum_{d=1}^{n} ICU$$

where F = weight of a destination area, d = destination area and ICU = Area (square metres) of industrial, commercial, public and private units derived from the Urban Atlas.

The third option is based on a different granularity, with the calculation based on individual ICU areas. The weight is equal to the area of each ICU and its position is represented by its centroid. Correlation analysis and the ArcGIS geoprocessing tool were used to perform calculations with the results stored in feature class datasets and ASCII grids. ASCII grids are transferred to maps (a,b) and compared. We used a cell-by-cell numerical comparison method based on the scaled absolute difference described by the following formula: abs (b – a) / max (abs (b – a)) (Visser, 2004). The result takes values from 0 to 1, where a value of 0 (red on the maps) indicates an area with no changes, while a value of 1 (green on the maps) indicates areas with major changes.

4. Results

The results were analysed in two ways. The first was a correlation analysis of all three measures (see Fig. 3). The correlations are quite high (0.958 – 0.988). The best correlation is between the two ICU area-based measures – a rather unsurprising result. This result suggests that the level of aggregation doesn't affect the results in a significant way and therefore the coarser granularity of the data seems to be quite acceptable on this scale – a result we didn't expect and which we explored in more detail.

The correlation between the census-based measure and the ICU area-based measures is also good, albeit with slightly lower levels of correlation. This is contrary to the conventional wisdom of significantly different job densities in ICU areas and appears to be a result of 'averaging' on a larger scale, suggesting that such a measure can be used as a proxy for more difficult to obtain data. The detailed view of the plots reveals some likely patterns – lines of points with probably slightly better correlation than the overall trend (e.g. the higher concentration of points with lower levels of accessibility with seemingly different or the line of points above the ellipse of best correlation in the plot of Census data jobs vs. individual ICU areas measures).

Category	**< 3.5% < 2°**	**3.5–5% 2–2.86°**	**5–6.5% 2.86–3.72°**	**6.5–8% 3.72–4.57°**	**8–15% 4.57–8.53°**
Motorway	120	100	80	n.a.	n.a.
Speedway (separated from other road traffic, mostly 4 lanes)	110	100	80	n.a.	n.a.
National road (category I)	80	80	70	60	n.a.
Other main roads (category II) + Other roads (category III)	70	70	70	60	50

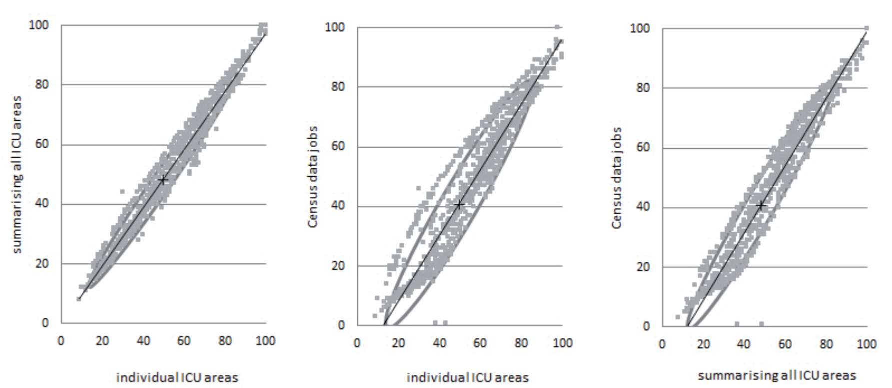
Tab. 2: Speed limit (km/h) is recalculated by slope for each type of road
Source: Authors' elaboration based on Franke et al., 2017

Category	**The speed in km/h when calculate the curvilinearity (grads)**		
	< 249	250–315	> 316
Motorway	130	110	100
Speedway (separated from other road traffic, mostly four lanes)	120	100	80
National road (category I)	100	90	80
Other main roads (category II)	90	80	70
Other roads (category III)	80	80	70

Tab. 3: The speed in km/h when calculate the curvilinearity (grads). Source: Franke et al., 2017

The map comparison results are somewhat different due to the patterns revealed and the differences between the strength of the patterns and the level of correlation. The patterns revealed in the map comparison between the measures of Census data jobs and aggregated (summarising) ICU areas show interregional differences without patterns at the local level (see Fig. 4). This result supports the findings of a good correlation between the two measures and reveals the effects of the different economic bases of the regions (e.g. a larger share of low-density manufacturing in the areas of Pilsen, Kolín and Pardubice compared to a larger share of office-related jobs in Prague).

The patterns of comparison of both ICU-based measures with different levels of aggregation are very different and contradict the superficial conclusions of high correlation of results and confirm the importance of the modifiable areal unit problem on the accessibility measures (see Fig. 5).



N		1225	
r	individual ICU areas	summarising all ICU areas	Census data jobs
individual ICU areas	1.000	0.988	0.958
summarising all ICU areas	0.988	1.000	0.974
Census data jobs	0.958	0.974	1.000

Fig. 3: Multivariate box plots of the results
Source: Authors' elaboration

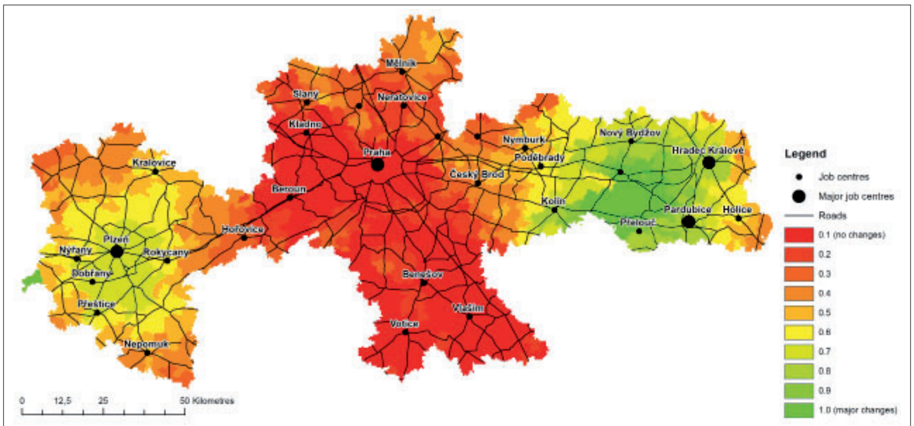


Fig. 4: Cell-by-cell numerical comparison method between attractiveness maps with weight parameters represented by Census data jobs (a) and summarising ICU areas (b). Source: Authors' elaboration

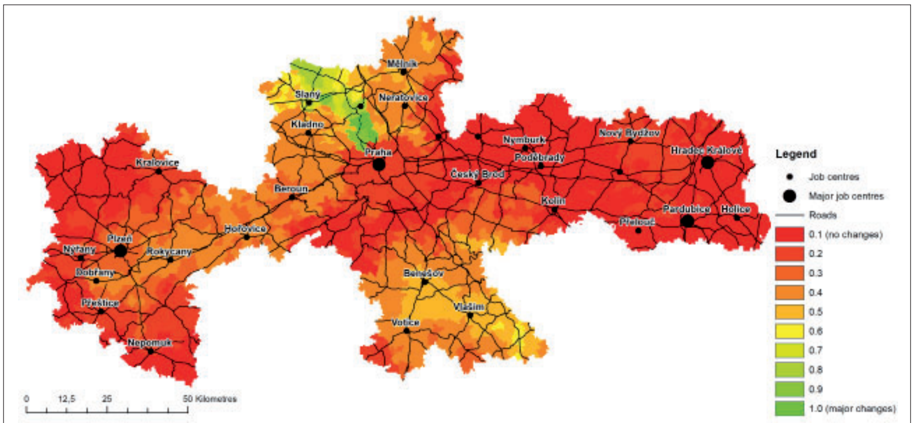


Fig. 5: Cell-by-cell numerical comparison method between attractiveness maps with weight parameters represented by summarising ICU areas (a) and individual ICU areas (b). Source: Authors' elaboration

The Prague hinterland is the most affected area due to the largest change in granularity in the Prague area. Two areas are strongly affected: the Kladno–Slaný area north of Prague and the Benešov–Votice area south of Prague. The former area has higher accessibility values if aggregation doesn't take place, as it is positively influenced by many commercial areas in the north-western part of the Prague city, including both high-density (e.g. retail centres, offices and universities) and low-density (e.g. warehousing and logistics in the vicinity of Prague Airport) areas. This effect is accentuated by the relative distance from the centre of Prague (definition point) and its complicated accessibility due to the lack of high-speed roads. The latter area around the towns of Benešov and Votice is an opposite case. Most of the ICU areas in the south-eastern part of Prague lie just beyond the boundaries of the Municipality of Prague along the D1 motorway and are therefore not affected by the aggregation of areas. The areas within the boundaries of the Municipality of Prague are either residential (large socialist-era housing estates) or rather small (mostly high-density office and retail areas). On the other hand, the aggregation reduces the relative distance to the areas on the NW edge of the Municipality of Prague and therefore the aggregation actually inflates the accessibility measures in this area.

The results outside the Prague city region are less affected. This is particularly the case in the east-west direction, with minor change in both the Pilsen and Kolín–Hradec Králové–Pardubice areas. There are two possible reasons for this: on the one hand, the decreasing relative size of the aggregation change at longer distances including decreasing importance of tangential road infrastructure and, on the other hand, the relatively balanced distribution of ICU areas in this direction within the municipality of Prague.

5. Discussion

The analysis of potential accessibility in the Czech Republic is often based on data at the level of municipalities (e.g. Maier et al., 2010) using counts of jobs due to its high validity of what is measured. The definition of the study area can be another source of error, as opportunities beyond the borders of the study area are often neglected. This leads to an underestimation or overestimation of the potential accessibility of the area. Our analysis has revealed that in some areas the effect of low validity related to spatial aggregation can be stronger than the effect of switch from high-validity jobs data to lower-validity data on jobs-related land-use areas.

The usual scales of place-based accessibility analysis depend mostly on the purpose of the analysis and the type of accessibility considered. The smallest scale is represented by analyses of pedestrian and other non-motorised accessibility (Lund, 2003; Iacono et al., 2010). The other extreme is represented by analyses carried out at continental or even global level, often with global city regions and countries as the finest unit of granularity. Examples of this scale include the analysis of the impact of TENs on potential population accessibility in Europe (Vickerman et al., 1999), modelling of the impact of the Channel Tunnel (Rohr & Williams, 1994), or more abstract accessibility-like analyses of global city networks (e.g. Derudder et al., 2013).

Our analysis is based on an accessibility indicator based on car and fits between the two extremes. While this seems to be an appropriate choice for an exploratory study it can underestimate the effects studied. For example, the accessibility by public transit is strongly affected by precise location of individual stops, in particular railway stations, and even small changes in jobs location can result in substantial changes in commuting time and corresponding distance decay factors.

While the results support the importance of the modifiable areal unit problem, the simplifications of the transport component probably underestimate it. It is mainly the omitted congestion of the inner city area – a factor that can significantly change the time distances in this area and thus increase the importance of the error caused by the modifiable areal unit problem.

6. Conclusions

The results are rather ambiguous. The high correlation between the results suggests that the data sets are good substitutes. The results of the map evaluation are more ambiguous and partly contradict the results of the simple correlation. Substitution seems to be possible as long as the scale of spatial change (e.g. changes in transport infrastructure) remains rather intra-regional and the level of aggregation remains similar.

The comparison of the three weight measures suggests that the accessibility of jobs in the central part of the Prague area is less dependent on the availability of extensive areas for commercial land-use. This can be explained by both the higher intensity of use of expensive land and the higher proportion of jobs in high-density mixed-use areas (e.g. services). The methods based on Census data jobs and aggregated ICU areas show similar results. The calculation of Census data jobs relies on the CZSO's census data, which are limited in scope and available only for the Czech Republic. On the other hand, the aggregation of all ICU areas is based on extensive data, widely applicable at EU level, but limited in the territory for which data are available (Urban Atlas).

The map comparison shows an important role of spatial aggregation on the results and their validity, especially in the Prague hinterland where the effect of different levels of aggregation is higher than the difference between the use of jobs vs. jobs-related land-use areas. The map-based comparison of both levels of aggregation of ICU areas contradicts the results of simple statistical analysis and reveals strong spatial patterns of differences.

The presented study provides only an exploratory analysis of the possibility of changing from data with high validity of what is measured to data with low validity of what is measured but higher validity related to spatial aggregation. Future research should address both the possibility to increase the validity of analysis based on land-use data and also explore different types of accessibility analysis and its sensitivity to modifiable areal unit problem and related validity. The results also show the usefulness of map analysis as a check on non-spatial statistical measures, as it can reveal strong spatial patterns at small scales that are hidden in the overall trends.

References:

Ahlfeldt, G. M. (2007). If Alonso Was Right: Accessibility as Determinant for Attractiveness of Urban Location. Hamburg Contemporary Economic Discussions, (12), 1–36. http://www.hced.uni-hamburg.de/WorkingPapers/HCED_12.pdf

Alonso, W. (1964). Location and Land Use: Towards a General Theory of Land Rent. Cambridge: Harvard University Press.

Boarnet, M., & Crane, R. (2001). The influence of land use on travel behavior: Specification and estimation strategies. Transportation Research Part A Policy and Practice, 35(9), 823–845. https://doi.org/10.1016/S0965-8564(00)00019-7

Chatman, D. G., & Nolan, R. B. (2011). Do Public Transport Improvements Increase Agglomeration Economies? A Review of Literature and an Agenda for Research. Transport Reviews, 31(6), 725–742. https://doi.org/10.1080/01441647.2011.587908

CZSO (2011). 2011 Census Results – Open Data. https://csu.gov.cz/otevrena_data_pro_vysledky_scitani_lidu_domu_a_bytu_2011_sldb_2011

CZSO (2021). 2021 Census Results – Open Data. https://csu.gov.cz/produkty/vysledky-scitani-2021-otevrena-data

CZSO (2021). Veřejná databáze. Praha. SLDB 2021. https://vdb.czso.cz/vdbvo2/faces/cs/index.jsf?page=profil-uzemi&uzemiprofil=34055&u=__VUZEMI__43__554782#

de la Barra, T., Pérez, B., & Vera, N. (1984). TRANUS-J: putting large models into small computers. Environment and Planning B: Planning and Design, 11(1), 87–101. https://doi.org/10.1068/b110087

Derudder, B., Taylor P. J., Hoyler M., Ni, P., Liu, X., Zhao, M., Shen, W., & Witlox, F. (2013). Measurement and interpretation of connectivity of Chinese cities in world city network, 2010. Chinese Geographical Science, 23(3), 261–273. https://core.ac.uk/download/pdf/55756574.pdf

de Vries, J. J., Nijkamp, P., & Rietveld, P. (2009). Exponential or power distance-decay for commuting? An alternative specification. Environment and Planning A, 41(2), 461–480. https://doi.org/10.1068/a39369

Echenique, M. H. (1994). Urban and regional studies at the Martin Centre: Its Origins, its Present, its Future. Environment and Planning B: Planning and Design, 21(5), 517–533. https://doi.org/10.1068/b210517

Echenique, M. H. (2011). Land use/transport models and economic assessment. Research in Transportation Economics, 31(1), 45–54. https://doi.org/10.1016/j.retrec.2010.11.007

El-Geneidy, A., & Levinson, D. (2006). Access to Destinations: Development of Accessibility Measures. University of Minnesota.

Erlebach, M., Tomáš, M., & Tonev, P. (2016). A functional interaction approach to the definition of meso regions: The case of the Czech Republic. Moravian Geographical Reports, 24(2), 37–46. https://doi.org/10.1515/mgr-2016-0009

Franke, D., Vorel, J., & Peltan, T. (2017). Job accessibility modelling in Prague Functional Urban Area. In Proceedings of the 2017 Smart City Symposium Prague (SCSP), Prague, The Czech Republic, 25–26 May 2017.

Geurs, K. T., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. Journal of Transport Geography, 12(2), 127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and Planning B: Planning and Design, 37(4), 682–703. https://doi.org/10.1068/b35097

Hansen, S. H. (2009). Analysing the Role of Accessibility in Contemporary Urban Development. In O. Gervasi, D. Taniar, B. Murgante, A. Laganà, Y. Mun, & M. L. Gavrilova (Eds.), Computational Science and Its Applications – ICCSA 2009. ICCSA 2009. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02454-2_27

Hanson, S., & Giuliano, G. (Eds.) (2004). The Geography of Urban Transportation. Third Edition, London: The Guilford Press.

Hewko, J., Smoyer-Tomic, K. E., & Hodgson, M. J. (2002). Measuring neighbourhood spatial accessibility to urban amenities: Does aggregation error matter? Environment and Planning A, 34(7), 1185–1206. https://doi.org/10.1068/a34171

Iacono, M., Krizek, K. J., & El-Geneidy, A. (2010). Measuring non-motorized accessibility: issues, alternatives, and execution. Journal of Transport Geography, 18(1), 133–140. https://doi.org/10.1016/j.jtrangeo.2009.02.002

IRP (2020). Chudé město pražské? Veřejné finance Prahy v českém a mezinárodním kontextu. Institut plánování a rozvoje hlavního města Prahy. https://iprpraha.cz/uploads/assets/5694_IPR_Chude_mesto_prazske_sazba_16_korektura.pdf

Kwan, M. P., & Weber, J. (2008). Scale and accessibility: Implications for the analysis of land use–travel interaction. Applied Geography, 28(2), 110–123. https://doi.org/10.1016/j.apgeog.2007.07.002

Logan, T. M., Williams, T. G., Nisbet, A. J., Liberman, K., Zuo, C. T., & Guikema, S. D. (2017). Evaluating urban accessibility: leveraging open-source data and analytics to overcome existing limitations. Environment and Planning B: Urban Analytics and City Science, 46(5). https://doi.org/10.1177/2399808317736528

Lopane, F. D., Kalantzi, E., Milton, R., & Batty, M. (2023). A land-use transport-interaction framework for large scale strategic urban modeling. Computers, Environment and Urban Systems, 104: 102007. https://doi.org/10.1016/j.compenvurbsys.2023.102007

Lowry, I. S. (1964). A Model of Metropolis, Santa Monica, CA: The RAND Corporation.

Lund, H. (2003). Testing the claims of new urbanism: Local access, pedestrian travel and neighboring behaviors. Journal of the American Planning Association, 69(4), 414–429. https://doi.org/10.1080/01944360308976328

Maier, K., Drda, F., Mulíček O., & Sýkora L. (2007). Dopravní dostupnost funkčních městských regionů a urbanizovaných zón v České republice. Urbanismus a územní rozvoj, 10(3), 75–80.

Maier, K., Mulíček, O., & Franke, D. (2010). Vývoj regionalizace a vliv infrastruktur na atraktivitu území České republiky. Urbanismus a územní rozvoj, 8(5), 71–81.

Martens, V., Lavalle, C., Batista e Silva, F., & Poelman, H. (2013). Population estimation for the urban atlas polygons. Joint Research Centre, Institute for Environment and Sustainability. https://data.europa.eu/doi/10.2788/54791

Micek, O., Feranec, J., & Štych, P. (2020). Land Use/Land Cover Data of the Urban Atlas and the Cadastre of Real Estate: An Evaluation Study in the Prague Metropolitan Region. Land, 9(5). https://doi.org/10.3390/land9050153

Modelistica (2011). TRANUS General description. https://modelistica.com.mx/files/GeneralDescriptionTranus.pdf

Modelistica (2012). Mathematical description of TRANUS. https://modelistica.com.mx/files/MathDescription.pdf

MoF (2024). Percentage share of individual municipalities in the shares of the national gross value added tax and income tax revenue. https://www.mfcr.cz/cs/kontrola-a-regulace/legislativa/legislativni-dokumenty/2024/vyhlaska-c-250-2024-sb-o-procentnim-podilu-jednotl-56947

Morris, J. M., Dumble, P. L. & Wigan, M. R. (1979). Accessibility indicators for transport planning. Transportation Research Part A: General, 13(2), 91–109. https://doi.org/10.1016/0191-2607(79)90012-8

Musil, J. (1966). Sociology of Urban Redevelopment Areas: A Study from Czechoslovakia. International Review of Community Development, 213–238.

Neis, P., Zielstra, D., & Zipf, A. (2011). The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. Future Internet, 4(4), 1–21. https://doi.org/10.3390/fi4010001

Novotný, V. (2011). Towards the Regional Acupuncture: How Far Is Too Far Away. Ph.D. Thesis, Faculty of life sciences, Czech University of Life Sciences in Prague.

Nyamai, D. N., & Schramm, S. (2023). Accessibility, mobility, and spatial justice in Nairobi, Kenya. Journal of Urban Affairs, 45(3), 367–389. https://doi.org/10.1080/07352166.2022.2071284

Openshaw, S. (1983). The Modifiable Areal Unit Problem, Norwich: Geo Books.

OSM (2011). OpenStreetMaps data. https://download.geofabrik.de/europe/czech-republic.html

Pazúr, R., Feranec, J., Štych, P., Kopecká, M., & Holman, L. (2017). Changes of urbanised landscape identified and assessed by the Urban Atlas data: Case study of Prague and Bratislava. Land Use Policy, 61, 135–146. https://doi.org/10.1016/j.landusepol.2016.11.022

Rohr, C., & Williams, I. N. (1994). Modelling the regional economic impacts of the Channel Tunnel. Environment and Planning B: Planning and Design, 21(5), 555–567. https://doi.org/10.1068/b210555

Rosik, P., Puławska-Obiedowska, S., & Goliszek, S. (2021). Public transport accessibility to upper secondary schools measured by the potential quotient: The case of Kraków. Moravian Geographical Reports, 29(1), 15–26. https://doi.org/10.2478/mgr-2021-0002

Spiekermann, K., & Wegener, M. (2007). European regional accessibility and the spatial impacts of transport energy price increases. In The accessibility and attractiveness of rural and landlocked areas: sustainable transport and services of general interest. European spatial planning and landscape, 87. https://www.spiekermann-wegener.de/pub/pdf/CEMAT_Accessibility_2007.pdf

UrbanAtlas (2012). Urban Atlas Land Cover/Land Use 2012 (vector). https://land.copernicus.eu/en/products/urban-atlas/urban-atlas-2012

Vickerman, R., Spiekermann, K., & Wegener, M. (1999). Accessibility and Economic Development in Europe. Regional Studies, 33(1), 1–15. https://doi.org/10.1080/00343409950118878

Visser, H. (2004). The MAP COMPARISON KIT: Methods, Software and Applications. https://www.researchgate.net/publication/27451992_The_MAP_COMPARISON_KIT_methods_software_and_applications von Thünen, J. H. (1826). Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie, oder Untersuchungen über den Einfluss, den die Getreidepreise, der Reichtum des Bodens und die Abgaben auf Ackerbau ausüben. Perthes.

Waddell, P. (2002). UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning. Journal of the American Planning Association, 68(3), 297–314. https://doi.org/10.1080/01944360208976274

Wee, B. (2022). Accessibility and equity: A conceptual framework and research agenda. Journal of Transport Geography, 104, 103421. https://doi.org/10.1016/j.jtrangeo.2022.103421

Please cite this article as:

Franke, D., Peltan, T., & Novotný, V. (2024). How data validity and granularity affect jobs accessibility: A case study from the Czech Republic. Moravian Geographical Reports, 32(4), 258–267. https://doi.org/10.2478/mgr-2024-0021