



Munich Personal RePEc Archive

# **Accessibility Models Based On the Gravity Analogy: In Theory and Practice**

Tóth, Géza and Kincses, Áron

Hungarian Central Statistical Office

July 2015

Online at <https://mpra.ub.uni-muenchen.de/73952/>

MPRA Paper No. 73952, posted 24 Sep 2016 11:15 UTC

GÉZA TÓTH<sup>a)</sup> – ÁRON KINCSES<sup>b)</sup>

## Accessibility Models Based On the Gravity Analogy: In Theory and Practice

### Abstract

The most commonly applied types of accessibility models are based on the gravity analogy. In these models, researchers use different types of resistance factors, but they rarely give any elaborate explanation for their choice of a specific type of factor in their research. Another problem with this kind of analysis is that in many cases, the authors do not describe precisely how they determine the constants for a line of calculations in a given model. Thus, the results cannot be fully accepted since they cannot be reproduced by the reader. Finally, we consider it to be yet another huge problem that the results of the models are rarely compared to the real (for example traffic) parameters, therefore, it is also impossible to detect what would happen if the researcher used a different model. In this study, we tried to line up the most commonly used models, and by enlightening the resistance factors, examine their possible usage and their boundaries through exact Hungarian examples.

*Keywords:* accessibility models, gravity analogy, resistance factors.

### Introduction

The models based on the gravity analogy are certainly the most commonly used type of models in the professional literature. (See for example: Stewart 1947, Hansen 1959, Ingram 1971, Vickerman 1974, Harris 1954, Huff 1963, Keeble et al. 1988, Dalvi–Martin 1976, Linneker–Spence 1991, Spence–Linneker 1994, Geertman–Ritsema van Eck 1995, Bruinsma–Rietveld 1998, Brunton–Richardson 1998, Kwan 1998, Levinson 1998, Smith–Gibb 1993, Gutiérrez 2001, Scheurer–Curtis 2007, etc.). These models attempt to take into consideration the impacts affecting the behaviour of the passenger. (Therefore, no matter which destination the passenger chooses, the possibility of any destination will be included in the model). One of the most common types of gravity analogy-based models is the potential model. The potential models – after forming disjoint, totally covered areal divisions – estimate the accessibility potentials of the area *i.*, regarding the other areas, from which those with the smaller mass or those who fall further have less of an impact (Rich 1980, Geertman–van Eck 1995).

The goal of this study is to present the background of the calculation of the models' constants. With the resulting constants, other accessibility models can be calculated, and

a) Hungarian Central Statistical Office, H-1024 Budapest, Keleti Károly utca 5-7., Hungary. E-mail: geza.toth@ksh.hu

b) Hungarian Central Statistical Office, H-1024 Budapest, Keleti Károly utca 5-7., Hungary. E-mail: aron.kincses@ksh.hu

then the results can be compared to the real flow data. We would like to demonstrate what kind of methods are used when selecting a given model, and how much it can affect the results derived from them. In this study, depending on their conditions, we offer suggestions on the applicability of the resistance factors.

### Material and methods

In this study, the definition of accessibility is: “Accessibility must refer to the role taken by the usage of an area and traffic system in a society that allows individuals and groups of individuals to participate in activities taking place in different locations.” (Geurs–van Wee 2004, p128). In this case – according to the definition – accessibility should be measured as a data numericizing the social space, which shows the density of the spatial connections of traffic systems and usage of lands in some areas. For the calculations, the actual areas must be simplified and modelled. This paper tries to describe the real spatial interfering impacts using the data resulting from the models.

The general potential formula of the potential models is the following:

Equation #1:

$$A_i = \sum_j D_j \cdot F(c_{ij}), \quad (1)$$

where  $A_i$  is the accessibility of the area;  $D_j$  is the mass of the area derived from  $i$ ;  $c_{ij}$  is the general cost of transportation between  $i$  and  $j$ ;  $F(c_{ij})$  is the resistance factor (function).

The earliest antecedent of this study is the gravity model by Hansen (1959). Hansen (1959, p78) claimed that accessibility is “the generalization of the connections of the population, reaching across distances”. The accessibility potentials of accessible destinations are strongly connected to the interaction of masses based on the gravity models.

The general formula of the Hansen-model:

Equation #2:

$$A_i = \sum_j \frac{W_j}{f(c_{ij}; \beta)}, \quad (2)$$

where  $A_i$  is the accessibility of the area;  $W_j$  is the mass accessible by the passengers, regardless of whether they actually want to access it or not;  $f(c_{ij}; \beta)$  is the resistance factor (the general marking of the resistance factor suggests that  $\beta$  can be part of an exponentiation as well as a multiplication, depending on the choice of the researcher);  $c_{ij}$  is a variable defining the cost of transportation between points  $i$  and  $j$ ; and  $\beta$  is a chosen constant.

The Hansen-model very much relied on the equations of gravity used in physics, which is also reflected in the fact that the constant in the equation, according to the physical deduction of the model, is bound to be a square number:  $\beta=2$ . (See: Calvo–Pueyo Campos–Jover Yuste, 1992).

Equation #3:

$$A_i = \sum_j \frac{W_j}{c_{ij}^{\beta}}, \quad (3)$$

The gravity and potential models are connected in many ways. What the two models have in common is that the size of a possible interaction between two settlements or locations is inversely proportional to the distance between them. Another similarity is that in the given settlement, any person (or other item) generates the same amount of interaction. Therefore, the size of the interaction between the two settlements is directly proportional to the size of the settlement, which is equivalent to chosen “mass”. Thus with the growth of the mass of the examined items, the extent of the interaction grows as well.

The problem with choosing the constant ( $\beta$ ) occurred many times in previous scientific studies, since in sociological analogies, it is not essential to stick with the squared powers used in the laws of gravity in physics. If the constant is larger than one, that means the researcher emphasizes the distances more. There are some types of analysis though where they model such infrastructural methods (for example, intercity trains), which can operate at their best in medium-length distances, in which case, the scale of the constant is one. (Martín–Gutiérrez–Roman 1999, Capineri 1996). These types of models have the advantage of being easy to comprehend and calculate. The measurements that consider each potential destination reflect the impacts affecting the behaviour of the passengers. Another great advantage is that with these measurements, the accessible locations can be distinguished from each other, meaning that their different values can be taken into account for the calculations. However, these models have a disadvantage in that they are unable to manage the passengers who have multiple traveling preferences. The results and the differences between the calculated numbers are also hard to interpret.

### *Choosing the Resistance Factor*

The usage of the distance dependence in social geography is mainly induced by the fact that the insularity in the area impedes the cooperation of different areas; hence, it should be numericized. The easiest cases are naturally the ones with geographical distances. Regarding the data on accessibility, only the travel time, distance or cost of travelling by some means of transportation should be taken into account. The distance to be covered between two given points is called the resistance factor.

The most notable difference between the adaptation of the potential model of accessibility and potential model of physics is that, unlike in physics, in sociology space is not continuous, but “discrete”. The social-economic forms (such as settlements or towns) tend to be concentrated in one distinctive point in space, and their “masses” are also concentrated in that spot. Since these “points of masses” do not fill the space, the potential value of a point – which obviously is altered by the impact of other points – in a limited part in the space (for example a country) could hardly be measured. (Tagai 2007). The mass points are concentrated in space to different extents; this induces different types of potential areas that might result in the distances between the points – and therefore the resistance factors – being represented by different functions. Very different resistance factors are required for dominance of small settlement structures versus large settlements, which affects the index exponent. Therefore, the equations of the resistance factor differ in the examinations of different areas, levels, or examinations of the same level, but not the same number of mass points.

Different forms of the resistance factor appear in the accessibility studies. In the case of models using boundaries, linear resistance factors are applied, or only destinations falling within a certain distance, time or cost are taken into account. Among models examining all accessible destinations and routes, the choice of the resistance varies considerably. The models distinguish the distances between the certain “masses” in their calculations. In the various approaches, the reciprocal or the exponentiated distance is applied. (See for example Hansen 1959, Davidson 1977, Fotheringham et. al. 2000, El-Geneidy–Levinson 2006.) The most commonly used types of models are the ones using a linear resistance factor (in the equation of the potential: in the denominator, the distance is raised to the first power). In these cases, no mathematical modifications are applied to the access time or the cost. However, this is not an unbreakable law, so in the models based on the gravity analogy, other exponents may occur. They numericize the probability of reaching the accessible destinations in the model. Essentially, researchers apply the exponential resistance factor for this purpose (Wilson 1971, Dalvi–Martin 1976, Martin–Dalvi 1976, Song 1996, Simma–Vritic–Axhausen 2001, Schürmann–Spiekermann–Wegener 1997, ESPON 2007, Papa–Coppola 2012). Other known models using resistance factors are the Gaussian ((Ingram 1971, Guy 1983) and the logistical ones (Bewley–Fiebig 1988, Hilbers–Veroen 1993). There are examples of other amendments to certain basic types, but this paper does not deal with them in detail (Reggiani, Bucci, Russo 2011).

Researchers divided the units of the access matrix (the units of the cost or duration of the passage between any examined parts of the area) into intervals (Simma–Axhausen 2003). What they found, is that the relationship between the frequencies and the average travel time/cost could best be described with an exponential regression function. These models are based on the presumption that within the examined area, with growth of the distance/travel time/cost, the probability of the frequency decreases exponentially, which is likely to affect the number of the potential destinations. According to this, the most subservient equation is:

Equation #4:

$$e^{-\beta c_{ij}}, \quad (4)$$

where  $c_{ij}$  is the travelling expenses (time) between the points,  $\beta$  is the constant.  $\beta$  is the constant of the examined spatial layout, which should be determined separately for each spatial structure. The reason why this is necessary, is that the connection between the different levels and in the case of examining different targets, the frequencies, and the average travel time/costs can be represented by various types of functions. The role of this constant is to make the connection between the contribution of the potential of the specific parts of space and the whole space per se. (Later, the problem of choosing the constant in detail will be discussed.)

When studying certain spatial structures, in exponential regression research, it is preferable to obtain an even more precise fit so we can determine the possibility of reaching certain destinations. To achieve this, it is vital to apply a Box-Cox (1964) transformation, which melds the residuals of the regression (making them homoscedastic), directing them to the normal distribution. Regarding the  $\epsilon_i$  mistakes, there is the presumption that their expected value is 0 and their deviations in divided groups are equal. This is the so-called *homoscedastic* case. If the mistakes in the measurement change in line with the variable X

(homoscedastic case), the occurring huge deviations (squared) give a disproportionately inaccurate range, thus inaccurate parameters as well; therefore, the results of regression or other models are not consistent with reality. When the homoscedastic condition is met, all the points of the regression line are straight, or the hyperplane residuals with the same deviation are located.

Equation #5:

$$\text{var}(\varepsilon_i) = \sigma^2 \quad \forall \quad i \in N^+ \quad \text{-re} \quad (5)$$

The Box-Cox transformation changes the values but not the order. For the usage of the Box-Cox resistance factor, the study by Willigers, Floor and van Wee (2007) gives an excellent example.

Ingram (1971) found that compared to the actual data, the values of some of the transformed resistance factors, distancing the origin, decrease too fast. He proposed the modified Gaussian resistance factor, which shows a slow decrease near the origin, and the extent of the decrease is smaller than that of the exponential and squared resistance factors. The smoothing nature of the Gaussian resistance function (it is convexo-concave) makes it apt to explore the spatial phenomena and – for instance – the migration of the population. (Grasland–Mathian–Vincent 2000) Equation #6:

$$f(d_{ij}) = 100 \cdot e^{-\frac{d^2}{u}} \quad (6)$$

In probability theory and statistics, the log-logistic distribution (in economics, “Fisk-distribution”) shows the probability of a continuous, not negative variable. It is used in areas where the probability of the variable is high initially, but then gradually decreases. The logistic distribution is a probability distribution of a random variable, the logarithm of which has a logistic distribution. The log-logistic models are based on the logistic distribution. The log-logistic distribution near the expected value is symmetric, although it can be described with a larger deviation than the log-normal one (since its expected value is calculated).

Equation #7:

$$f(d_{ij}) = 1 + e^{a+b \cdot \ln d} \quad (7)$$

### ***The Self-potential***

The professional publications on potential models have been studying the notion of “self-potential” (see among others Frost–Spence 1995, Bruinsma–Rietveld 1998). The significance of this is that, within examined space, the extent of the place-dependent potential is not only defined by the sizes and distances of the masses around it, but also depends on how big of a force field it can generate around itself. In the studies of potential, it is vital to distinguish inner from outer potential (Nemes Nagy 1998, 2005). This severance comes from the distinction of the specifically examined area and the force of the space, outside that area, affecting it. So the potential of a given point is calculated from the sum, the self-, inner and outer potentials.

Self-potential should also be taken into consideration in the cases of accessibility studies. When calculating the self-potential of an area, it is presumed that the transportation

between two areas is not the only condition increasing accessibility, but also the transportation within areas or settlement. Thus, we conclude that a certain product or service does not necessarily have to be transported to another area if it can be sold within the given area. Neglecting the role of the self-potential might be misleading – for example – when examining the level of settlements. Naturally, in these cases, the accessibility of central settlements of the agglomeration or group of settlements would be much lower than the other settlements of the agglomeration.

When defining the self-potential – like when examining other types of potential – the calculations are derived from the size of the given area (preferably, not the administrative size, but the “clear” one). According to the generally used methods, the area is regarded as a circle, and then the radius for a given unit that is considered proportionate to the distances on roads within certain settlements is calculated; thus, it is also called “self-distance”. This type of distance is used in models calculating geographical distances, whereas with those applying distances based on the road network, the distance can be converted by an average velocity/cost formula, which can then be placed in the equation. The difference between the various methods used for calculating the self-potential is that to what extent and in what way the radius is weighted (Tagai 2007). In this way, by examining it from different aspects, the role of the centre point can be further highlighted, or it can be better fitted to the surface of potential.

Tagai's study claims that there are some researchers who use distances equivalent to the radius since its length is close to the average distance within the area.

The most commonly used method is calculating a third of the radius as the self-potential. This is also connected to the probability of the mass distribution. Others, however, came to different conclusions after similar calculations; they believe two-thirds of the radius is the most suitable as the self-derived approximate value of distance.

Modifying the radius in any way is arbitrary, and consequently, hard to back up objectively. Therefore, in this study, it is always the whole radius that is considered when calculating the self-potential. Naturally, if actual data about the settlement is available, the various assumptions can be compared to the facts; although, in this case, it is also true that some models fit some settlements better.

### ***The Total Potential***

In the accessibility studies, usually the examined space is chosen so that it is larger than the narrow examination area, and in this way, the impact of the outer potential can be ignored. In this case, the space-dependent accessibility potential is calculated from the sum of the inner and self-potential:

Equation #8:

$$\Sigma A_i = SA_i + BA_i \quad (8)$$

where  $\Sigma A_i$  is the total potential of area  $i$ ,  $SA_i$  is the own potential of the spatial unit, and  $BA_i$  is the inner potential within the area analysed. There are some theories that take into account the potential outside the area, the “outer potential”, however, within this study, on the level of smaller areas, the outer potential cannot be calculated due to methodological reasons, and hence they are not included. Although this affects the result of the examination

to some extent, it does not prevent us from drawing up basic connections. The effect can mostly be noticed on the peripheries; the work did not deal with the models that take the competition into account (Weibull 1976, Knox 1978, Van Wee Annema Hagoort, 2001, Joseph Bantock 1982, Fotheringham 1982).

### Calculation

In the professional literature on accessibility, the method for calculating the constant is not often discussed. Thus, readers can only interpret the results or recreate the research with considerable difficulty. The next segment covers the theoretical background determining these constants in detail and the way they are calculated.

By definition, accessibility numericizes the possibility of approaching a location from the aspects of the passengers, regarding their (household, business) options and goals; furthermore, the services available at the desired destination, and relationships connected to the space. The models based on the gravity analogy – in our opinion – conclude from some elements of the definition that the number of the potential journeys from point A to point B is determined by four factors: the desired destination's mass and distance; the examined area's spatial structure and the role of random probability. Three out of these four factors can be modelled, whereas conclusions can be drawn from the forth, but they are hard to predict.

It would not be true to state that the number of journeys between two points half as far away from each other is twice as many; or – in line with that – a point three times the mass is visited three times as many times. Yet the modelling can be done. The goal is to numericize the probability of the potential journey. In this case, the potential is the function of three factors: the desired destination's mass, distance and spatial structure. The first two factors are specific to each potential model; however, in the case of models using a constant, the latter should be described with a factor. With the functions used (exponential, exponential modified by the Box-Cox transformation, Gauss-function), the role of the spatial structure has different impacts on the value of the potential; they infiltrate in different ways depending on the geographical distance. The constants of the mathematical functions in the resistance factors are defined by the frequency of the accessibility distances appearing in the examined space (Simma–Axhausen 2003 p. 184).

The paper will now focus on the exact calculations. The first unit in the potential's formula is the self-potential while the rest is the inner potential's impact on the self-potential. The formula can be applied to any optional area – for example a settlement, a county, or a region. The following equation determines the value of the constant  $\beta$ .

The value of the place-dependent potential in point  $j$  of the space:

Equation #9:

$$A_i = W_i \cdot e^{-\beta \cdot c_{ii}} + \sum_{i \neq j} W_j \cdot e^{-\beta \cdot c_{ij}} = \frac{W_i}{e^{\beta \cdot c_{ii}}} + \sum_{i \neq j} \frac{W_j}{e^{\beta \cdot c_{ij}}} \quad (9)$$

where  $A_i$  is the accessibility of the area;  $W_i$  and  $W_j$  are the masses – in this case, population – belonging to the appropriate level;  $c_{ij}$  is the time needed to get to area  $i$  from area  $j$ , on the



road network, measured in minutes. Lastly,  $\beta$  is the constant of the examined spatial structure, which should be determined separately in each spatial structure.

It should be noted that the current definition of the potential implies there is a linear superposition between the units, thus there is no interaction between the impacts; they do not enhance or lessen each other, they simply add up. Looking for analogies, the gravitational, magnetic or electric fields are much similar. The definition does not take into account the interference described in string theory as in quantum physics.

The potential journeys on the road network between all possible pairs of all the 175 districts are investigated as follows.

The data has been laid out in a  $175 \times 175$  matrix, aligned with the travel times (in minutes). The values are divided into intervals, and avoid placing too few occurrences in one interval, yet forming a sufficient number of intervals, because that would hinder the examination of the distribution of the values. In the study, the access times were divided into 50 equal intervals.

Take the set containing intervals that contain the durations of the journeys between the districts in minutes:

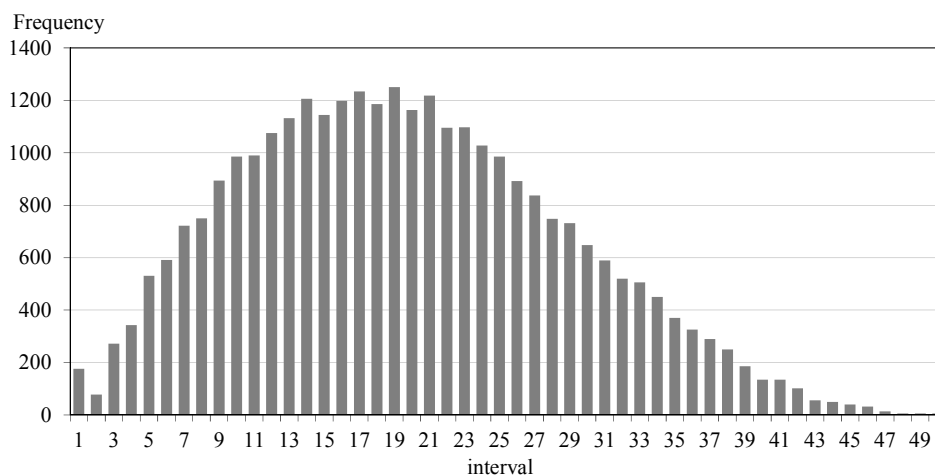
$$\forall \chi \in \Phi : \chi \in (0; 429,17) .$$

The 429.17 minutes result is the largest unit of the  $175 \times 175$  matrix; thus, this is the longest distance between points of the district on the road network, measured in minutes. In this examination, theoretical access times were used for the calculations, meaning that the length of time was only affected by the speed limits and not any other conditions.

Divide the interval into 50 equal pieces. Interval number “ $i$ ” contains the times ( $i \cdot 8.58$ ;  $(i+1) \cdot 8.58$ ); where  $i = 1, 2, \dots, 50$ , thus dividing the maximum mid-district distances of 429 minutes into 50 equal pieces results in 8.58-minute-long intervals.

Figure 1

*The frequency of the accessibility time*



Source: own compilation.

A function-like connection should be found between the frequency of the intervals and the average travel times. The function is needed for modelling the frequency of journeys

with the increase in the distances, which technically translates to the probability of the journey. Theoretically, the goal is to find the function most suitable for the frequency since that would result in the most accurate potential model. However, human behaviour (in this case, travelling) can barely be modelled – as can be seen in many cases of models based on a physical analogy. It is not certain at all that the most suitable model will give the most accurate results. Moreover, it may occur that the frequency of the intervals fit the given model very poorly, yet the model using the constant calculated from it gives the closest to the actual results.

At first glance, the polynomial approach may look like the obvious choice, but it would make the results difficult to interpret. Applying it would be necessary due to the fluctuation shown by the frequency. Naturally, this could vary in each sample. Further problems may occur when trying to interpret the rate of the frequency together with the increase of the distances. Probably the most obvious case is that of the exponential function since here the frequency decreases exponentially as the distances grow. As a result, the version that finds the exponential connection between the travel times and the mean values of the intervals is reviewed first.

### *The Exponential Resistance Factor*

Equation #10:

$$V_l \approx e^{-\bar{c}_l} \quad l=1,2,\dots,50, \quad (10)$$

We assume that the duration of the journeys among the districts and the mean values of the intervals in minutes are proportional to each other. The constant  $\beta$  makes the exact connection between the average length of time and the frequency.

Equation #11:

$$v_l = e^{-\beta \bar{c}_l} \quad l=1,2,\dots,50, \quad (11)$$

where  $v$  is the frequencies and stands for the average duration. So the presence of the exponential connection mentioned above should be checked for each examination; the value of the constant should also be calculated each time.

The equation (11) implies a regression relationship, and we are looking for the constant  $\beta$  that best fits the equivalences overall. After rearranging the equation, we get:

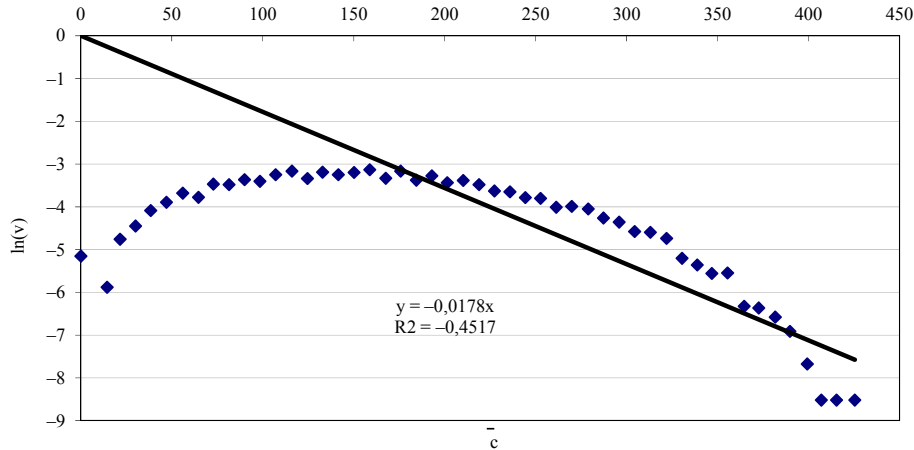
Equation #12:

$$\ln v_l = -\beta c_l \quad (12)$$

Depicting the  $e$  (a natural number) based logarithm of the frequency in the function of the average travel times, the value of  $\beta$  can be determined by linear regression. In the study, the normalizing criterion of the joint line crossing the origin is required. Thus, the normalized frequency belonging to the average time of 0 is 1. The calculated value of  $\beta$  is 0.0178, with 45.17% reliability.

Figure 2

The  $n$  based logarithm of frequency with the function of average durations



Source: own compilation.

The method described above is just the first, approximate solution. This schema ignores the statistical errors occurring during the division-based access times, since, during this process, only the outer multiplications and subtractions among the groups were taken into account. See the following:

Take  $c_{ij}$ , (in this case  $ij$  does not mean the duration of travel between  $i$  and  $j$ !) as the value number “ $j$ ” from the interval number “ $i$ ”. ( $i=1, 2, 3, \dots, 50$ ;  $j=1, 2, \dots, p_i$ ). Now any optional element of the set  $\Phi$  can be described as such:

Equation #13:

$$c_{ij} = \bar{c}_i + e_{ij} \tag{13}$$

where  $\bar{c}_i$  is the average of the interval  $i$ ;

$$\bar{c}_i = \frac{1}{p_i} \sum_{j=1}^{p_i} c_{ij}, \text{ and } e_{ij} \text{ stands for the errors and residues of the groups.}$$

Take a fully squared value:

Equation #14–15:

$$\begin{aligned} Q_{all}^2 &= \sum_{i=1}^l \sum_{j=1}^{p_i} (c_{ij} - \bar{c}_{..})^2 = \sum_{i=1}^l \sum_{j=1}^{p_i} (c_{ij} - \bar{c}_i + \bar{c}_i - \bar{c}_{..})^2 = \\ &= \sum_{i=1}^l \sum_{j=1}^{p_i} (c_{ij} - \bar{c}_i)^2 + \sum_{i=1}^l \sum_{j=1}^{p_i} (\bar{c}_i - \bar{c}_{..})^2 + \\ &= 2 \sum_{i=1}^l \sum_{j=1}^{p_i} (c_{ij} - \bar{c}_i)(\bar{c}_i - \bar{c}_{..}) \end{aligned} \tag{14}$$

Applying  $2 \sum_{i=1}^l \sum_{j=1}^{p_i} (c_{ij} - \bar{c}_i)(\bar{c}_i - \bar{c}_{..}) = 0$ , since  $\sum_{j=1}^{p_i} (c_{ij} - \bar{c}_i) = \sum_{j=1}^{p_i} c_{ij} - p_i \bar{c}_i = 0$ ; (15)

Therefore  $Q_{all}^2 = Q_{inner}^2 + Q_{outer}^2$ .

And where

Equation #16-18:

$$Q_{inner}^2 = \sum_{i=1}^l \sum_{j=1}^{p_i} (c_{ij} - \bar{c}_i)^2; \quad (16)$$

$$Q_{outer}^2 = \sum_{i=1}^l \sum_{j=1}^{p_i} (\bar{c}_i - \bar{c}_{..})^2; \quad (17)$$

$$\bar{c}_{..} = \frac{1}{n} \sum_{i=1}^l \sum_{j=1}^{p_i} c_{ij} \quad (18)$$

This applies to its deviation as well:

$$\sigma_{all}^2 = \sigma_{inner}^2 + \sigma_{outer}^2 \quad (19)$$

where  $\sigma_{inner}^2$  is the deviation within the intervals squared;  $\sigma_{outer}^2$  is the deviation in between the intervals squared.

Therefore, the phenomenon described the way it was defined in the first part of the study is only accurate if  $e_{ij}$  of the equation  $C_{ij} = \bar{c}_i + e_{ij}$  has the following features:

1. Their expected value in every interval is zero.
2. Within the intervals, they have the same deviation. In other words, they are homoscedastic.

If the conditions are not met, the results will differ depending on the classification, thus will not be reliable enough to base realistic predictions on them. In the case regarding the districts – described above – we find that the residues are usually large, and the two requirements are only approximately fulfilled. Using the first method would not be a huge mistake by any means, but there is a method in statistics to eliminate even that little flaw. Namely, it is the Box-Cox transformation, which secures the random occurrence of residues. With this method, the data that best meets the two conditions above and fits a given spatial layout can be calculated.

**The Box-Cox Resistance Factor**

Two forms of the Box-Cox transformation are known.

Equation #20-21:

$$1. \quad c_{ij}^{transformed} = \begin{cases} \frac{c_{ij}^\lambda - 1}{\lambda}, \lambda \neq 0 \\ \ln(c_{ij}), \lambda = 0 \end{cases} \quad (20)$$

$$2. \quad c_{ij}^{transformed} = c_{ij}^\lambda \quad (21)$$

Because of the rearrangement,  $C_{ij}$  must be larger than 0; this condition has to be met by every access time measured in minutes.

With the restriction of  $\lambda > 0$ , it can be achieved in both cases that the transformation is relation invariant, meaning that this transformation might change the values, but not the order. Both definitions lead to the same conclusion. We chose the latter. Therefore, we are looking for the value of  $\lambda$  that best suits the random distribution of differences. For the investigations, we used the software SAS 8.2, more specifically, the proc transreg. This program calculates the transformed duration times with all possible values of  $\lambda$  and very little class differences, and the log-likelihood functions belonging to these intervals. The estimate of the maximum log-likelihood function is technically an estimate of a point in which we consider those parameters that have the maximum combined partition function of the examination vector. The log-likelihood function:

Equation #22:

$$\log L(c_{ij}, f_k, \lambda) = F(c_{ij}, \lambda) \cdot F(f_k, \lambda) \quad (22)$$

Estimating the maximum log-likelihood function has the advantage of being asymptotically effective, furthermore, if it cannot be given in closed-form, it can be examined by numeric maximization (this applies to this case). In other words, the population parameters can be estimated by the value where the log-likelihood function reaches its maximum; thus, the most probable outcome is a realized sample when taking a sample.

In the district example,  $\lambda = 1.07085$  was calculated for  $\lambda$  belonging to the maximum output of the function. This is how the transformation constant was arrived at, which, by transforming the access times, makes the residues of the groups the most independent. Overall, with this method, the information resulting from the group formations with transformed access times, is more accurate.

In the case of a transformed variable, if the boundaries of the 50 intervals are transformed in the same way, we get an identical exponential curve. In this way, we can redefine the average travel times of the groups, and determine the value of  $\beta$  by regression, based on the previous, exponential calculation. In our example, regarding already transformed pairs of travel times,  $\beta = 0.0119$ . Therefore, we can conclude that constant  $\beta$  can be determined. This has to be done separately for each examination or spatial distribution. For the first estimate, we find that  $\beta = 0.0178$ , while after further investigation, it was proven that the Box-Cox transformation carried out with the power of  $\lambda = 1.0785$ , allows a more accurate analysis. In this case,  $\beta = 0.0119$ .

### ***The Gaussian Resistance Factor***

According to figure 3, when analysing the distribution of frequencies, the following relationship between the frequencies and average access times can be presumed:

Equation #23:

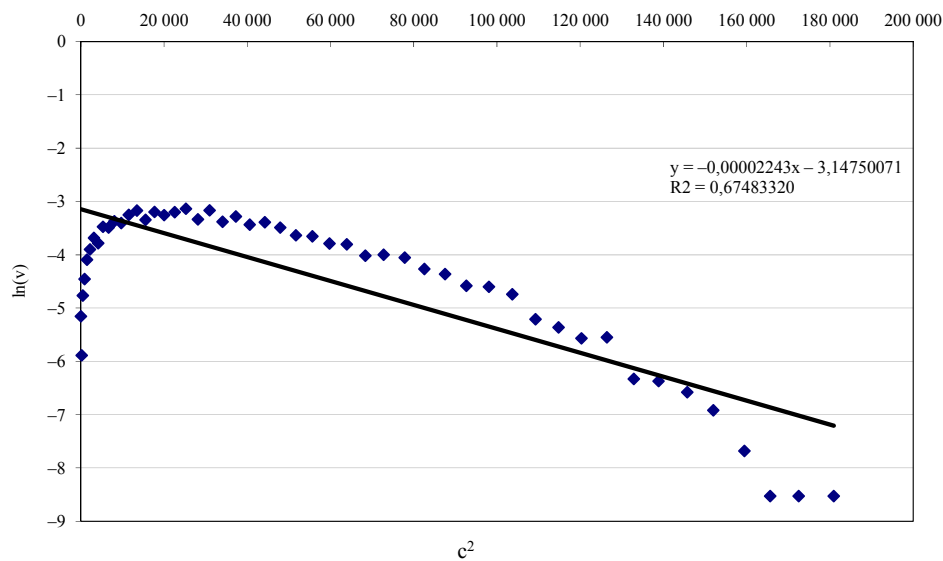
$$v \approx w \cdot e^{\frac{-c_i^2}{u}} \Rightarrow \quad (23)$$

$$\ln(v) \approx \ln(w) + \frac{1}{u}(-c_i^2)$$

where  $v$  is the frequencies,  $c$  is the average access times, and  $w$  and  $u$  are the constants. The goal is to calculate the constants that best fit the existing data. For this purpose, we represented the logarithm of the frequency of the elements belonging to the 50 intervals formed based on the access times between the capitals of the districts in the function of the average durations of the squared intervals.

Figure 3

*The logarithm of frequencies and the average durations of the intervals squared*



Source: own compilation.

From which  $w = 0.04295936$ ,  $u = 44583.1476$ ; with 0.67 accuracy. The Gauss model better fits that in an exponential function.

***The Log-logistic Resistance Factor***

In this case, we predict the following relationship between the frequency and the average duration:

Equation #24:

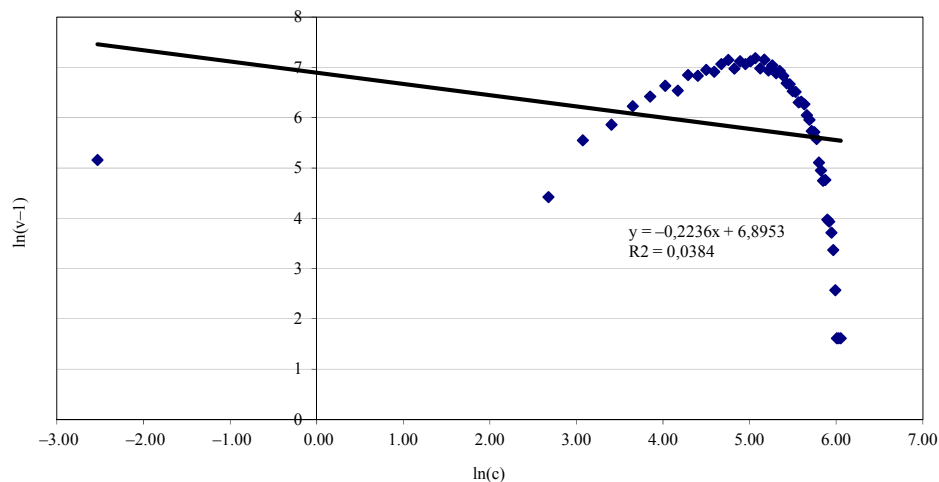
$$v \approx 1 + e^{a+b \ln c_i} \Rightarrow \tag{24}$$

$$\ln(v - 1) \approx a + b \ln c_i$$

where  $v$  stands for the frequencies,  $c$  stands for the average durations;  $a$  and  $b$  are constants.

Figure 4

*The relationship between the frequency and the average duration in a log-logistic case*



Source: own compilation.

From which:  $b = -0.2236$ ,  $a = 6.8953$ .

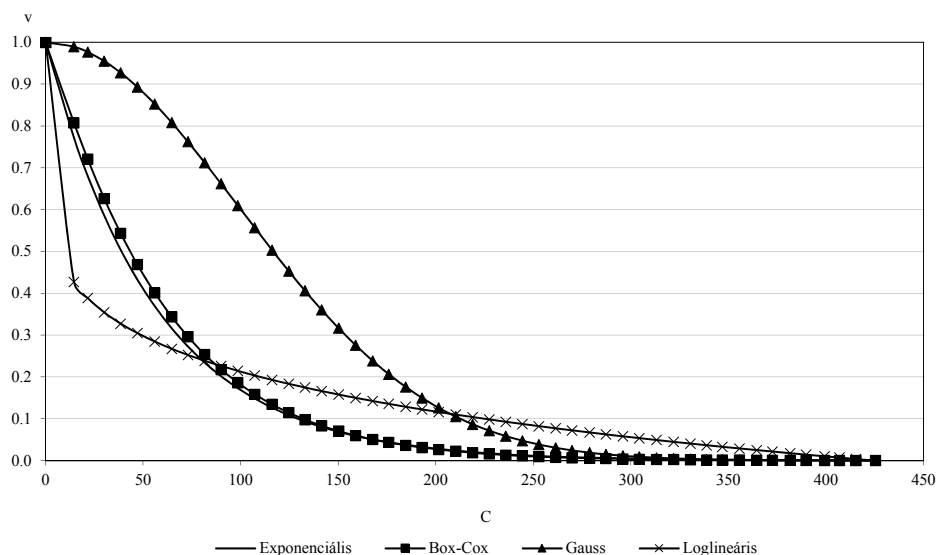
It should be noted that  $R^2$  is smallest in this example, yet this method – as can be seen later – may lead to accurate results.

For comparing the different types of resistance factors, we redid these calculations. The recalculation of the Gaussian and log-logistic model was needed to fit the others, meaning that in case of  $c = 0$ , the frequency is 1 (or 100%). (According to these calculations,  $R^2$  is decreasing; therefore, the data resulting from them cannot be used further, and are only represented in figure 5.)

By comparing the various resistance factors, the differences between them can be studied. The access times are measured in minutes with the distances constructed according to the specific models. Noticeably, the Gaussian model best works for the medium distances, the log-logistic approach for the very large or very small while the exponential method for the smaller distances. Theoretically, according to this, for the studies within settlements or districts, the log-logistic approach should be used; for country-level studies the exponential; for a European study the Gaussian, and for a global examination once again the log-logistic model. Yet – as will be shown later – it is not certain that these approaches provide the most accurate results in the mentioned areas.

Figure 5

*The relationship between the resistance factors of the models based on the gravity analogy*



Source: own compilation.

## Results

The study has so far focused on the structures of the different models, ignoring the question of how realistically the potential structure formed by basic elements and relationships describes the space (based on geometric interpretations). We have not yet dealt with how comparable the volume of traffic on the public roads is with the data derived from the models. So can the models' conclusions be applied to the real social space? During the research, we only found studies in the professional literature that compare the results of different models to each other (see De Montis–Caschili–Chessa 2011); we could not find one that contrasts them with the traffic. Thus, this essay is unique in the professional literature.

The traffic data measured by the Hungarian Public Road Not-for-profit Inc. keeps track of the annual average daily traffic (AADT) for the cross-section of certain roads. (The latest nationwide traffic measuring included the cross sections of over 4500 roads.) These measurements are carried out by taking samples. This method allows researchers to calculate the average daily traffic – if they are aware of the time-specific fluctuation of traffic – from relatively little data (smaller sample, briefer counting) with sufficient accuracy and reliability.

The nationwide cross-section measurements take place at a large number of junctions and deviations throughout the year, on five different occasions, from 6 am to 6 pm.



Most of the annual measurements last for three days. A rotating method is used so that annually, measurements are only carried out in about 20% of the country, and then the data from the previous years are modified to fit the new results.

From the measured data, by simple averaging, and then multiplying by factors bearing the time-specific characteristics of the traffic at a given time of the day ( $a_x$ ), a given day ( $b_i$ ) or month ( $c_i$ ), the annual average daily traffic can be calculated.

Equation #25:

$$\dot{A}NF = \frac{1}{n} * \sum_{i=1}^n g_x * a_x * b_i * c_i \quad (25)$$

where  $n$  is the number of the days counted;  $g_x$  is the traffic measured within the  $x$  hours;  $a_x$  is the time-of-day factor (the volume of traffic at a specific time of day in proportion to the whole 24 hours);  $b_i$  is the daily factor (a multiplier for the given day of the week that modifies the daily traffic to the monthly average);  $c_i$  is the monthly factor (a multiplier for a given month of the year to give the yearly average).

The results of the cross-section traffic count are expanded to the “validity sections” according to a convention of the experts. The district-level AADT data was provided by the Hungarian Public Road Not-for-profit Inc.

Table 1

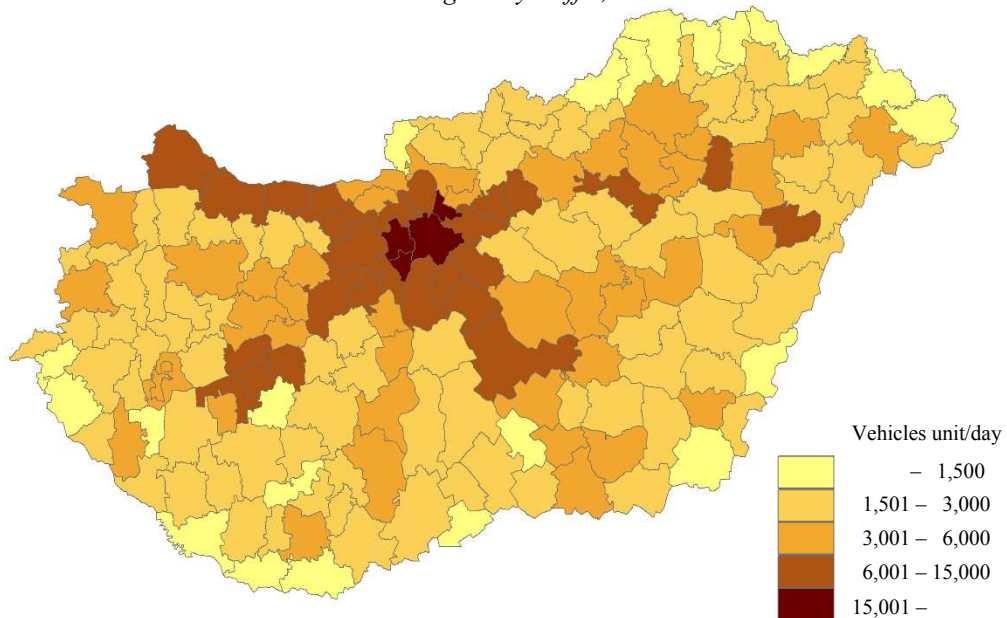
*The dimensions of the study*

Dimension	Notes
Source	In the study, the accessibility is calculated from the aspect of each individual; also, we interpret but do not distinguish certain social groups or the varying destinations of the different passengers.
Goal	We numericize the aspired destinations by its population and income.
Resistance	In this case, the areal resistance factor is the theoretical access time between given district capitals on the public road network, measured in minutes. The resistance factor applied can be linear, squared, exponential, Box-Coy, Gaussian, or log-logistic.
Restrictions	When travelling on the roads between two districts, the maximum achievable speed on that specific type of road is the speed limit.
Boundaries	Hungary's borders mean the boundaries of the examined area. Although it is an undeniable fact that the Hungarian potentials are affected by destinations outside the country since there was no sufficiently detailed map of the road network of other countries available, we had to neglect their impact.
Means of transportation	We did not distinguish personal or freight traffic for this study.
Modus	The calculations regarding public roads were based on unimodal accessibility.
Areal level	The study mainly focuses on the district level, LAU 1.
Diversity	The study's main goal is to model the differences in accessibility in Hungary.
Dynamics	In the study, the population, income, and public network as of January 1 <sup>st</sup> : 2004, 2008 and 2012, is included.

Source: own compilation.

Figure 6

*Annual average daily traffic, 2012*



Source: own compilation.

We compared 2004, 2008 and 2012 data with various potential models. As a mass factor, both the income and population was included in the calculations. (For details of these models see: Tóth–Kincses 2007). The dimensions of the study can be seen in Table #1.

**Models Based on the Gravity Analogy**

*General gravity analogy*

The examined models are the following:

Equation #26-31:

$$c_i \cdot \frac{W_i}{c_i} \cdot \sum_j \frac{W_j}{c_j} \tag{26}$$

$$c_2 \cdot \frac{W_i}{c_i} \cdot \sum_j \frac{W_j}{c_j} \tag{27}$$

$$c_3 \cdot \frac{W_i}{e^{\beta c_i}} \cdot \sum_j \frac{W_j}{e^{\beta c_j}} \tag{28}$$

$$c_4 \cdot \frac{W_i}{e^{\left(\frac{c_i-1}{\lambda}\right)}} \cdot \sum_j \frac{W_j}{e^{\left(\frac{c_j-1}{\lambda}\right)}} \tag{29}$$

$$c_5 = \frac{W_i}{p * e^{-\frac{c_{ij}^2}{u}}} + \sum_j \frac{W_j}{p * e^{-\frac{c_{ji}^2}{u}}} \tag{30}$$

$$c_6 = \frac{W_i}{1 + e^{a + b \ln c_{ij}}} + \sum_j \frac{W_j}{1 + e^{a + b \ln c_{ji}}} \tag{31}$$

where  $C_{1-6}$  is the accessibility of area  $i$ ,  $W_i$  is the mass of the district,  $W_j$  is the “mass” of the desired destination,  $c_{ij}$  and  $c_{ji}$  are the access times, while  $\beta, \lambda, p, u$  and  $a, b$  are constants. (Equation 26 is using the linear model, 27 is using the squared, 28 the exponential, 29 the Box-Cox, 30 the Gaussian and 31 the log-logistic model.)

As Tables 2 and 3 show us, the traffic data can be better fitted together with the income numbers rather than with the population numbers, although the difference is not significant. The experiments on a district level suggest that the most suitable potential models are those using the log-logistic factor; it should be noted that if other types of division had been applied, the results would have been the same. While calculating with the log-logistic resistance factor, the parity of the duration frequencies was quite weak. Despite this, the approach provided the most accurate model. This demonstrates that it is appropriate to use more than one type of model for the calculations; we cannot draw conclusions only from parity.

The reason the log-logistic function is favourable here is that it is the one that can estimate the differences of the areal traffic, with the help of the explanatory variable. In case of the areal differences of traffic in Hungary, the most important aspect is to find the function that can not only take the basic areal differences into account, but is able to estimate, with the smallest residue possible, the traffic value of the capital city, Budapest. If Budapest’s impact were not so outstanding or if there was a district with a similarly large volume of traffic in Hungary, the calculations carried out with log-logistic function might not give the most accurate results.

Table 2

*The parity of the models based on the gravity analogy using population masses, compared to the district AADT statistics (R<sup>2</sup>)*

Years	c <sub>1</sub>	c <sub>2</sub>	c <sub>3</sub>	c <sub>4</sub>	c <sub>5</sub>	c <sub>6</sub>
2004	0.43	0.26	0.55	0.52	0.19	0.63
2008	0.45	0.45	0.56	0.52	0.13	0.69
2012	0.58	0.44	0.61	0.58	0.21	0.69

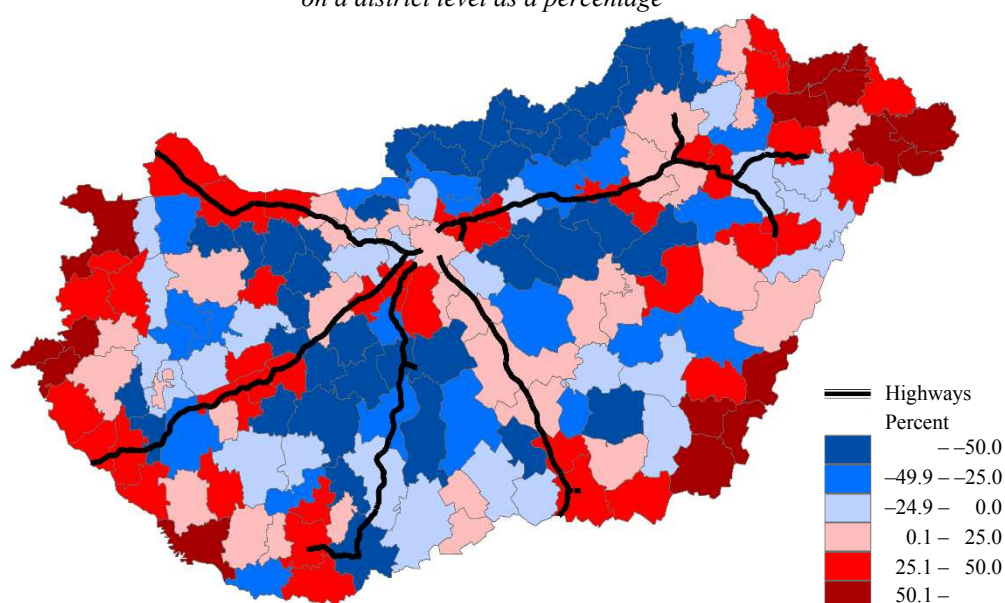
Table 3

*The parity of the models based on the gravity analogy using income masses, compared to the district AADT statistics (R<sup>2</sup>)*

Years	c <sub>1</sub>	c <sub>2</sub>	c <sub>3</sub>	c <sub>4</sub>	c <sub>5</sub>	c <sub>6</sub>
2004	0.42	0.24	0.56	0.53	0.18	0.73
2008	0.46	0.45	0.58	0.55	0.11	0.72
2012	0.58	0.45	0.60	0.57	0.22	0.69

Figure 7

*The 2012 average daily traffic; the potential model with the largest parity (c<sub>6</sub> – with log-logistic resistance factor); the difference of the 2012 estimated traffic on a district level as a percentage*



Source: own compilation.

In one-half of the districts, mainly the busiest ones that are near a highway, the traffic estimated by the models is less than the actual (these are marked red on the map). In the rest of the districts, located in the middle of the country, the volume is usually underestimated while districts near the border are once again overestimated.

## Results and Discussion

The methodological base for applying different types of resistance factor is to numerize the role of the spatial structure in the model. The possibility of the journeys – the potential – is determined by the mass of the desired destination, the distance and random chance. In this case, “spatial structure” refers to the size of the road network in the analysis, by the disruption of which, the distances/travel costs are modified with various functions.

The goal of choosing one resistance factor from the several available (squared, Box-Cox, Gaussian, exponential, log-logistic) is to differentiate the various relationships by the probability to be chosen. Out of the four factors of travel potential, the travel relationship is only one; the remaining three factors have further impacts on the probability of the choice. The spatial relationship can hardly be described mathematically. The choice of the resistance factor cannot be generalized according to the area or the extent of the analysis. It is essential – if it is all possible – to compare the calculations carried out with different resistance factors to find the most suitable model for the analysis.

\*\*\*

**Acknowledgement**

This paper is supported by the **János Bolyai Research Scholarship** of the Hungarian Academy of Sciences.

## REFERENCES

- Bewley, R.–Fiebig, D. G. (1988): A flexible logistic growth model with applications to telecommunications *International Journal of Forecasting* 4 (2): 177–192.
- Box, G. E. P.–Cox, D. R. (1964): An analysis of transformations *Journal of Royal Statistical Society, Series B* 26 (2): 211–246.
- Bruinsma, F.R.–Rietveld, P. (1998): The Accessibility of European Cities: Theoretical Framework and Comparison of Approaches *Environment and Planning A* 30 (3): 499–521.
- Brunton, P.J.–Richardson, A.J. (1998): A Cautionary Note on Zonal Aggregation and Accessibility. *Paper presented at the 77th Annual Meeting of Transportation Research Board*, Washington, DC.
- Calvo, P.–Pueyo Campos, A.–Jover Yuste, J. M. (1992): *Potenciales demográficos de España Atlas Nacional de España*. Tomo 14-b. Instituto Geográfico Nacional, Madrid.
- Capineri, C. (1996): From Networks to Regional Development: Representations of Italian Regional Disparities. *Paper Presented at the Nectar Euroconference*, 24–28 September 1996., Mons, Belgium.
- Dalvi, M. Q.–Martin, K. M. (1976): The measurement of accessibility: some preliminary result. *Transportation* 5 (1): 17–42.
- Davidson, K. B. (1977): Accessibility in transport/land-use modelling and assessment. *Environment and Planning A*. 9 (12): 1401–1416.
- De Montis, A.–Caschili, S.–Chessa, A. (2011): Spatial Complex Network Analysis and Accessibility Indicators: the Case of Municipal Commuting in Sardinia, Italy *European Journal of Transport and Infrastructure Research* 11 (4): 405–419.
- El-Geneidy, A. M.–Levinson, D. M. (2006): *Access to Destinations: Development of Accessibility Measures*. Department of Civil Engineering University of Minnesota, Minneapolis.
- ESPO (2007): *Update of Selected Potential Accessibility Indicators*. Final Report. Luxembourg.
- Fotheringham, A. S.–Brunsdon, C.–Charlton, M. (2000): *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage, London.
- Fotheringham, A. S. (1982): A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning A*. 15 (1): 15–36.
- Frost, M.E.–Spence, N.A. (1995): The Rediscovery of Accessibility and Economic Potential: The Critical Issue of Self-potential. *Environment and Planning A*. 27 (11): 1833–1848.
- Geertman, S.C.M.–van Eck, J.R.R. (1995): GIS and models of accessibility potential: an application in planning. *International Journal of Geographical Information Systems* 9 (1): 67–80.
- 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.
- Grasland, C.–Mathian, H.–Vincent, J. (2000): Multiscalar analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences. *Statistical Journal of the United Nations Economic Commission for Europe* 17 (2): 157–188.
- Guy, C. M. (1983): The assessment of access to local shopping opportunities: a comparison of accessibility measures. *Environment and Planning B: Planning and Design*. 10 (2): 219–238.
- Hansen, W.G. (1959): How Accessibility Shapes Land-Use. *Journal of the American Institute of Planners*. 25 (2): 73–76.
- Harris, C. D. (1954): The market as a factor in the localisation of industry in the United States. *Annals of the Association of American Geographers* 44 (4): 315–348.
- Hilbers, H. D.–Veroen, E. J. (1993): *Het beoordelen van de bereikbaarheid van lokaties. Definiëring, maatstaven, toepassing en beleidsimplicaties*. INRO-VVG 1993–09, TNO Inro, Delft.

- Ingram, D. R. (1971): The Concept of Accessibility: A Search for an Operational Form. *Regional Studies* 5 (2): 101–105.
- Joseph, A.E.–Bantock, P.R. (1982): Measuring potential physical accessibility to general practitioners in rural areas: a method and case study. *Social Science and Medicine* 16 (1): 85–90.
- Keeble, D.–Offord, J.–Walker, S. (1988): *Peripheral Regions in a Community of Twelve Member States*, Commission of the European Community, Luxembourg.
- Knox, P.L. (1978): The intraurban ecology of primary medical care: patterns of accessibility and their policy implications. *Environment and Planning A* 10 (4): 415–435.
- Kwan, M. P. (1998): Space–time and integral measures of individual accessibility: a comparative analysis using a point-based framework. *Geographical Analysis* 30 (3): 191–216.
- Levinson, D.M. (1998): Accessibility and the Journey to Work. *Journal of Transport Geography* 6 (1): 11–21.
- Linneker, B.J.–Spence, N.A. (1991): An Accessibility Analysis of the Impact of the M25 London Orbital Motorway on Britain. *Regional Studies* 26 (1): 31–47.
- Martín, J.C.–Gutiérrez, J.–Román, C. (2004): Data Envelopment Análisis (DEA) Index to measure the accessibility impacts of new infrastructure investments: the case of the High-Speed Train Corridor Madrid–Barcelona–French border. *Regional Studies* 38 (6): 697–712.
- Martin, K. M.–Dalvi, M. Q. (1976): The comparison of accessibility by public and private transport. *Traffic Engineering and Control* 17 (12): 509–513.
- Nemes Nagy, J. (1998): A fekvés szerepe a regionális tagoltságban. In: *Munkaerőpiac és regionalitás*, pp. 147–165., MTA KK KI, Budapest.
- Nemes Nagy, J. (ed) (2005): *Regionális elemzési módszerek*. Regionális Tudományi Tanulmányok 11. ELTE Regionális Földrajzi Tanszék–MTA–ELTE Regionális Tudományi Kutatócsoport, Budapest.
- Papa, E.–Coppola, P. (2012): Gravity-Based Accessibility measures for Integrated Transport–land Use Planning (GraBAM) In: Hull, A.–Silva, C.–Bertolini, L. (Eds.) *Accessibility Instruments for Planning Practice*, pp. 117–124. COST Office.
- Reggiani, A.–Bucci, P.–Russo, G. (2011): Accessibility and Impedance Forms: Empirical Applications to the German Commuting Network. *International Regional Science Review* 34 (2): 230–252.
- Rich, D. (1980): *Potential Models in Human Geography*. Concepts and Techniques in Modern Geography. 26. Geo Abstracts, Norwich.
- Scheurer, J.–Curtis, C. (2007): *Accessibility Measures: Overview and Practical Applications*. Department of Urban and Regional Planning, Curtin University, Perth.
- Schürmann, C.–Spiekermann, K.–Wegener, M. (1997): *Accessibility Indicators*. Berichte aus dem Institut für Raumplanung 39, IRPUD, Dortmund.
- Simma, A.–Axhausen, K.W. (2003): Interactions between travel behaviour, accessibility and personal characteristics: The case of the Upper Austria Region *European Journal of Transport and Infrastructure Research* 3 (2): 179–197.
- Simma, A.–Vrtic, M.–Axhausen, K. W. (2001): *Interactions of travel behaviour, accessibility and personal characteristics: The Case of Upper Austria, presentation*, European Transport Conference, Cambridge, September 2001.
- Smith, D.M.–Gibb, R.A. (1993): The Regional Impact of the Channel Tunnel. A Return to Potential Analysis. *Geoforum* 24 (2): 183–192.
- Song, S. (1996): Some Tests of Alternative Accessibility Measures: A Population Density Approach. *Land Economics* 72 (4): 474–482.
- Spence, N.–Linneker, B. (1994): Evolution of the motorway network and changing levels of accessibility in Great Britain. *Journal of Transport Geography* 2 (4): 247–264.
- Stewart, J.Q. (1947): Empirical mathematical rules concerning the distribution and equilibrium of population. *Geography Review* 37 (3): 461–485.
- Tagai, G. (2007): A potenciálmodell erényei és korlátai a társadalomkutatásban *Tér és Társadalom* 21 (1): 145–158.
- Tóth, G.–Kincses, Á. (2007): Elérhetőségi modellek. *Tér és Társadalom* 21 (3): 51–87.

- 
- van Wee, B.–Hagoort, M.–Annema, J.A. (2001): Accessibility measures with competition. *Journal of Transport Geography* 9 (3): 199–208.
- Vickerman, R.W. (1974): Accessibility, attraction, and potential: a review of some concepts and their use in determining mobility. *Environment and Planning A* 6 (6): 675–691.
- Weibull, J.W. (1976): An axiomatic approach to the measurement of accessibility. *Regional Science and Urban Economics* 6 (4): 357–379.
- Willigers, J.–Floor, H.–van Wee, B. (2007): Accessibility indicators for locational choice of offices: An application to the intraregional distributive effects of high-speed rail in the Netherlands. *Environment Planning. A* 39 (9): 2086–2098.
- Wilson, A. G. (1971): A family of spatial interaction models, and associated developments. *Environment and Planning A* 3 (1): 1–32.