Accessibility analysis for planning applications I: impedance functions

Anastasia Soukhov

The term **Accessibility** has many definitions. Within the context of transportation planning literature and practice, accessibility is often a location-based measure of *potential for interaction*. Specifically, it quantifies the potential a “population” has to reach “opportunities” in a given region based on their means of transportation. The people or activities at some origin in space and time are the “population”: they could be individuals employed at a type of job, children of a certain age group or other characteristic, or simply *all people* or some other type of opportunity-seeking activity (like a business) that reside at origins. The “opportunities” are the type of destinations that the “population” interacts with, and the definition of its selection is as critical and numerous as the selection of “population”. Further, modes (e.g., walking, transit), time of travel, quality of route taken, and quality of “opportunities” are among many factors that can be considered within accessibility measures.

The output from accessibility measures are typically a value or scaled score that is assigned to each spatial unit (e.g., a census tract, neighbourhood boundary, parcel, etc.). This output provides a snapshot of the relationship between land-use and transportation in the region: areas with high scores are relatively well-connected and are in proximity to plenty of opportunities while the opposite is true for areas with low accessibility. Accessibility analysis can be used by planners to identify priority areas for transportation and improvements in “opportunities”.

Needless to say, accessibility analysis can be an important component when planning for equitable transportation and service provision.

This blog post is the first of a series of posts that present accessibility analysis for planning applications. The series will walk readers through the components of accessibility analysis as well as its potential uses when planning for equity. This post explores how travel behaviour enters accessibility measures through the *impedance function*, the implications of travel behaviour assumptions and how one may select parameters for these assumptions. In the subsequent posts, I plan to discuss how these assumptions about travel behavior impact accessibility analysis outputs, different types of accessibility measures, and centering equity and justice conceptualizations in accessibility analysis.

## Counting opportunities based on travel behaviour assumptions

Many accessibility measures derive from the work of (Hansen 1959) represented in ([Equation 1](#eq-hansen-access)):

The accessibility score at each origin is a weighted sum of the number of opportunities at destinations , where and are a set of spatial units in a region. The weights in this summation are a function of the cost of travel , sometimes called a distance-decay function. reflects how the potential for interaction changes with the cost of travel between spatial units and , that is the origin and destination of a potential trip.

The cost in the distance-decay function can be distance, time, financial cost, or a combination of several factors. Since distance is not always the unit of travel cost, is also known more generally as an *impedance function* since the function models the impedance of travel. Generally, declines with growing travel cost (the impedance is greater), and so opportunities at destinations that are less costly to reach are more heavily weighted in the summation that yields . Conversely, opportunities that are costly to reach (i.e., they are *far* away in terms of travel cost) have values of that are close to or equal to zero, so a negligible amount of enters the summation.

In short, the impedance function allows the accessibility analyst to precisely define a measure of travel behavior: the relationship between the “population” at an origin and where they usually, want, or can go to reach “opportunities” at destinations.

From this perspective, the definition of the impedance function is incredibly important. Let’s go over commonly defined impedance functions in accessibility research and their impact on opportunity-counting (the summation of opportunities) at specific travel costs , namely:

* Binary ([Equation 2](#eq-binary-access))
* Uniform distribution ([Equation 3](#eq-uniform-imped))
* Exponential distribution ([Equation 4](#eq-exp-imped))
* Gamma distribution ([Equation 5](#eq-gamma-imped))

The **binary function** ([Equation 2](#eq-binary-access)) forms the basis of the cumulative opportunities measure approach (discussed in the next [post](**LINK%20TO%20POST2**)). The binary function is *binary* because it returns only two values, typically either 1 and 0. If the opportunity is reachable from to within some sort of travel cost threshold , it returns a 1 for that trip. Conversely, it returns 0 if the travel cost is above a certain threshold , meaning the opportunity exceeds the cost that people are willing to travel to reach it.

Threshold should be selected carefully to reflect the observed or assumed travel behavior for the situation of interest. For instance, assume the travel cost is in the units of car travel minutes. Does only counting the potential interaction opportunities for the population in a region accessing destinations within a 0 to 15 minute range make sense? If yes, then the threshold in this function would be 15. This means that only those opportunities that can be reached within 15 minutes from any given spatial unit will be counted. All other opportunities that are reachable beyond 15 minutes will be assigned a 0.

The impedance function can take other forms, such as the commonly used [*probability density functions*](https://en.wikipedia.org/wiki/Probability_density_function) (PDF): values can be interpreted as the *probability density* of a trip occurring for each value of travel cost . If probability density values are plotted on the y-axis for each travel cost along the x-axis, the probability of a trip occurring between a certain range of is the area under the curve. Important to note is that the area under a PDF always sums to 1, i.e., 100% probability that the trip between the minimum and maximum will occur.

The **uniform distribution** PDF looks very similar to the binary function, as it only returns one of two values. However, it also has the property of PDFs - the area under the curve for the range of is always 1. The general form for the uniform distribution PDF is shown in ([Equation 3](#eq-uniform-imped)). The parameters that the analyst chooses are and : these represent the maximum and minimum travel costs (i.e., the range) that describe the observed or assumed willingness to reach destinations. If the trip is of a travel cost that is within this range, it returns a value of . Outside of this range, the function returns a 0, so we are assuming the potential of the population to interact with those opportunities is zero.

However, analysts using a binary threshold must ask themselves: is it true that populations only travel to opportunities that are within a 15 minute drive? Is this 15 minute cut-off a fair assumption to make about their travel behaviour? Maybe it’s more accurate to assume that the probability of a trip does not strictly drop to *zero* beyond 15 minutes. In this case, it would be worth while considering other distributions.

Other types of functions are the **exponential distribution** and the **gamma distribution**. The theoretical form of these two PDFs are shown in ([Equation 4](#eq-exp-imped)) and ([Equation 5](#eq-gamma-imped)). The analyst must select parameters for these functions represented by (exponential) and and (gamma).

For the **exponential distribution**, the probability of a trip occurring is always highest at the lowest value of travel cost (e.g., a trip that has a travel cost of 1 has a higher probability density than a trip with a travel cost of 10). The mediates the rate of the exponential curve; specifically, the higher the parameter value, the higher the rate of travel cost decay. So at a value that is large, the majority of trips occur within a smaller range than if the was a smaller value. Though the exponential distribution is more complex than the uniform, it allows the analyst to model travel behaviour without having to select a binary threshold beyond which opportunities are no longer counted.

Consider another situation: if the probability of a trip occurring is not always highest at the lowest value of travel cost, then the **gamma distribution** can be considered. In fact, for the **gamma distribution**, the probability is often low at low costs, higher at mid-costs, and low again at high costs. The and parameters controls the rate and shape of the gamma curve. The higher the (gamma rate) parameter, the higher the probability of the majority of trips occurring within a low travel cost range. So at low (gamma rate) parameter values, the same probability is spread across a larger range of travel costs. For the (shape) parameter, the higher value, the higher the probability density of trips with a higher mean travel cost.

Values for both and are used in the gamma distribution, so it is more complex in formulation than the exponential. However, the gamma may be more useful in modelling specific travel behaviour. Namely, if the population’s travel behaviour is less likely to occur at short travel times, more likely at mid-range travel times, and less likely at long travel times, the gamma distribution can be calibrated to match this pattern. This form of travel behaviour can occur within observed home-to-work commutes from predominately single-use zoned regions: trips are less likely to occur at short travel times for a region (as a result of single-use residential zoning), are more likely at mid-range travel costs (commuting to a central business district), and less likely at long travel costs (few super-commuters). Representing this travel behaviour pattern cannot be accurately captured by the exponential distribution as short travel times *should* have low values of . Similarly, the use of the uniform distribution is inaccurate in this situation as it requires the analyst to select min. and max. travel cost thresholds such that the opportunities that the short- and long- travelling population potential interactions are not counted (i.e., returns value of 0).

These three PDF distribution forms are presented in order of increasingly complexity. As the complexity increases, the flexibility of explaining the travel behaviour also increases. I created an interactive [R Shiny Application](https://soukhova.shinyapps.io/Impedance-explained-shiny-app/) to experiment with the parameter values and help conceptualize what each distribution form may mean for travel behaviour assumptions by interpreting the “probability density of trip” (y-axis) at values of travel cost (x-axis) . Give it a try!

# An empirical example: calibrating a model that reflects travel behaviour

In all the impedance forms presented, the analyst must define parameters. A clever technique is to build a trip length distribution (TLD) using empirically observed origin-to-destination (OD) travel survey data. A TLD reflects observed travel patterns: specifically, how likely an observed trip of a certain travel cost is to occur for the population in a region of interest. Based on the TLD, we can select the best fitting theoretical PDF forms (e.g., uniform, exponential, gamma), fit the associated parameters (e.g., & $T\_{max$, , or and ) and use the calibrated theoretical PDF to carry the assumptions about the population’s travel behaviour into the accessibility calculation.

Here I demonstrate an overview of calibrating a PDF for a sample of empirical home-to-work travel flows taken from workers who live and work (full-time) within the City of Hamilton from the R data package [{TTS2016R}](https://soukhova.github.io/TTS2016R/). The flows are aggregated at level of traffic analysis zones (TAZ). This package contains a subset of home-to-work flows from the 2016 Transportation Tomorrow Survey (TTS) as well as road-network car travel times from TAZ centroids (calculated using [{r5r}](https://doi.org/10.32866/001c.21262)). {TTS2016R} is detailed in this publication (Soukhov and Páez 2023) and is freely available [here](https://soukhova.github.io/TTS2016R/).

The TLD for this empirical data is shown below ([Figure 1](#fig-TLD-empirical)):

|  |
| --- |
| Figure 1: Trip length distribution of home to full-time work trips (in estimated minutes by car) for the City of Hamilton. |

In our example, the y-axis is the probability density of a trip at a given travel cost in minutes of travel . It can be observed that the probability density of a trip is highest when travel time is around 11 minutes. It can also be seen that beyond the 30 min mark approximately, the rate of probability density drastically decreases. So, the probability of a trip of length 0 to 30 mins occurring is 95% (the area under the curve between these two x-value points is 0.95). Trips outside of this range make up the remaining probability.

Now let’s fit the parameters of the uniform, exponential, and gamma functions ([Equation 3](#eq-uniform-imped), [Equation 4](#eq-exp-imped), [Equation 5](#eq-gamma-imped)) as closely to the TLD captured in [Figure 1](#fig-TLD-empirical). The R package [{fitdistrplus}](https://cloud.r-project.org/web/packages/fitdistrplus/index.html) was used to generate parameters that best fit the TLD. The moment matching estimation (MME) fitting method and the Nelder-Mead direct optimization algorithm are used (Delignette-Muller and Dutang 2015). The default values for the parameters of the three functions are summarized as follows:

* : and is 0 and 29 mins, respectively.
* : (rate) is 0.07
* : (shape) is 3 and (rate) is 0.2

|  |
| --- |
| Figure 2: Trip length distribution (empirical) with fitted theoretical PDFs (coloured) of home to full-time work trips for the City of Hamilton. |

For curves shown in [Figure 2](#fig-TLD-all): the higher the , the higher the probability density of travelling to reach the opportunities at the destination.[[1]](#footnote-1)

The uniform impedance function (red), when implemented into an accessibility calculation, would assume that the population is indifferent to *changes* in travel cost. The population at an origin is assumed to either totally interact with an opportunity (if it’s a trip between 0 to 29 minutes - the thresholds) or not interact at all.

If the exponential (green) or gamma curve (blue) was implemented in an accessibility calculation, then the analyst is assuming the population is much more sensitive to changes in travel cost. However, the exponential and gamma functions are quite a different shape, so they depict a different response to the probability of traveling given a travel cost .

The exponential curve (green) is more intuitive to understand: the shorter the travel cost , the higher the value. However, when compared to the empirical curve (black) (i.e., the observed travel behaviour), we can see they do not closely match. Trip lengths that are 11 mins in length have the highest probability density of occurring and trips that are longer and shorter than this length occur less often and are assigned decreasing values. For these reasons, the gamma function (blue) provides a fit that is closest to the empirical curve at the cost of a more complex mathematical formulation.

The impedance function reflects significant assumptions about travel behaviour. The selection of the type of function and associated parameters reflects the impedance that populations face reaching opportunities. How the impedance function is used to explain accessibility is discussed in next [post](**LINK%20TO%20POST2**).

Again, feel free to explore the parameters interactively for the uniform, exponential and gamma distributions using the interactive Shiny R Application [here](https://soukhova.shinyapps.io/Impedance-explained-shiny-app/).

*The TLD used in this post is a subset of data from {TTS2016R}, the goodness-of-fit criteria and diagnostics from {fitdistrplus} are used for model parameter selection, plots are generated using {ggplot2}, and spatial objects are manipulated using {sf}, along with base {R} functions. Feel free to view all the code and text in this post (including the interactive plot) in the* [*GitHub repository*](https://github.com/soukhova/MJ-Accessibility-Blogs)

## References

Delignette-Muller, Marie Laure, and Christophe Dutang. 2015. “fitdistrplus: An R Package for Fitting Distributions.” *Journal of Statistical Software* 64 (4): 1–34. <https://doi.org/10.18637/jss.v064.i04>.

Hansen, Walter G. 1959. “How Accessibility Shapes Land Use.” *Journal of the American Institute of Planners* 25 (2): 73–76. <https://doi.org/10.1080/01944365908978307>.

Soukhov, Anastasia, and Antonio Páez. 2023. “TTS2016R: A Data Set to Study Population and Employment Patterns from the 2016 Transportation Tomorrow Survey in the Greater Golden Horseshoe Area, Ontario, Canada.” *Environment and Planning B: Urban Analytics and City Science*, January, 23998083221146781. <https://doi.org/10.1177/23998083221146781>.

1. Using {fitdistrplus}, the parameters in all theoretical functions were selected through an optimization algorithm that minimizes the differences between all possible parameter range(s) and the empirical function for each theoretical function. [↑](#footnote-ref-1)