

UWZGLĘDNIANIE NIEPEWNOŚCI W MODELACH POTENCJAŁÓW RUCHOTWÓRCZYCH¹

Lajos Kisgyörgy

PhD, Budapest University of Technology and Economics, Hungary, tel.: +36 1 463 1159, e-mail: kisgyorgy@uvt.bme.hu

Gergely Vasvari

MSc, Budapest University of Technology and Economics, Hungary, tel.: +36 1 463 1537, e-mail: vasvari@uvt.bme.hu

***Streszczenie.** W modelach sieci transportowych wiele zmiennych opisuje cechy fizyczne sieci. Integracja wynikowej niepewności jest tylko częściowa. Wiele tych zmiennych jest opartych na hipotezach i oszacowaniach. Metoda Monte Carlo jest efektywną narzędziem oceny scenariuszy, w których pewne zmienne nie mogą być opisane dokładnie. W artykule rozważano zastosowanie metody MC z zaproponowaną strukturą zmiennych losowych opisujących niepewność podczas generowania podróży. Powtarzanie obliczeń powoduje rosnącą dokładność wartości potoku ruchu na łuku i różnorodność funkcji rozkładu. Szczegóły analizy ilustrują trudności, które mogą pojawiać się podczas stosowania metody MC.*

***Słowa kluczowe:** metody Monte Carlo, modelowanie transportu, VISUM, symulacja stochastyczna*

1. Introduction

This paper establishes a connection between the Monte Carlo method and transportation modelling. To introduce the implementation, general outlines are addressed first, briefly discussing the characteristics of the MC method, and relevant methodologies of modelling.

Details of the analysis illustrate the difficulties which might rise when adopting the MC method. Results of a hypothetical scenario are presented in regard of link volumes, showing different random effects.

Further possibilities promise higher convergence rate, estimation of precision, and sensitivity analysis of the modelled transportation network in context of the implemented stochastic variables. Extending the set of these variables is also a possible path.

¹ Wkład autorów w publikację: Kisgyörgy L. 50%, Vasvari G. 50%

2. Uncertainty, risk, confidence

Measurements and estimations are usually about supporting decisions[1]. If the decision in question has significant consequence would it turn out wrong, then methods reducing uncertainty about it have a high value. The meaning of uncertainty is worth clarifying: uncertainty is the lack of complete certainty, which means the existence of more than one possibility, and that the “true” outcome is not known. Uncertainty could be measured by assigning a set of probabilities for the set of possible outcomes. By this approach risk is defined as a state of uncertainty where some of the possibilities involve a loss, catastrophe or other undesirable outcome. Risk is measured by a set of possibilities each with quantified probabilities and quantified losses.

One method to express uncertainty about a variable is to think of it as a range of probable values. In statistics a range that has specific probability to contain the correct value is called a confidence interval.

For any kind of model which – at least partially – aims for prediction coping with uncertainties may have a high value. However usual methods to predict the expected traffic volumes are strongly limited in this sense. The results of traffic forecast models depend greatly on the development of the future values of independent variables. And this development is unclear, we cannot tell the future values for sure. This uncertainty can have significant consequences hence it should be evaluated. One possible method to deal with uncertainty is stochastic simulation which describes the result with probability distribution.

3. Monte Carlo method

The Monte Carlo method is a powerful and practical empirical method to reduce uncertainty[2,3]. In general terms, Monte Carlo method always consists of a high number of repeated calculations, where the values of some input variables are determined randomly before each calculation[4]. After the whole process – often referred as Monte Carlo Simulation (MCS)[5]– results are evaluated either to describe the expected value and probability distribution of an unknown quantity or to estimate the results of complex equations where the analytical solution is not feasible. Hence the method requires the ability to describe the uncertainty of several variables – which would be generated according to this description – and also a deterministic model, or process to calculate dependable results for each repetition. Conventionally it is used when there is a need to account for the uncertainties in future values of variables affecting the outcome of a deterministic theory (prediction), or to assess the possible outcome of a complex event-tree (decision support, reliability analysis, risk management), and even to acquire the value of a complex procedure when the required precision can be set (computational physics).

The applied Monte Carlo method first assigns a probability density function (PDF) to the uncertain variables. The type of the distribution usually is normal, triangular, uniform or lognormal, based on the conditions surrounding the variable, the current state of knowledge and personal experience. Once the distributions are attributed, MCS takes randomly sampled values from them, to form one possible scenario and one possible result. After recording the outcome the process is repeated several times, giving the response of measurements for a range of possible scenarios. The measured variables are then complemented by a probability distribution that approximates the answer to the problem.

4. Transportation modelling

To clarify the terms of the research, it should first be noted that the demonstrated method was tested on a small size hypothetical transportation network model to explore implementation possibilities. Public transportation mode and complex demand modelling was not addressed in this model. The proposed method was tested with the basic 4-step modelling approach first, to assess its suitability. Therefore, the advanced topics will not be discussed here. Thorough calibration of the model was also omitted as the current goal was not to give reasonable prediction for a unique scenario but to explore the theoretical possibilities and benefits of the stochastic simulation.

The MC method requires the identification of deterministic processes and uncertain variables in the environment where the simulation would be applied. These are detailed in the following paragraphs.

4.1. *Deterministic features*

A network model[6] is a definite description of the real-life road network[7]. Movement restrictions, impedances, capacities all depend on road network characteristics. If these qualities changed over time that would mean a different model in the analysis.

Basic trip distribution[8] methods deliver a solution for splitting the previously generated traffic demand into O-D matrices. Distribution with the gravity model is an iterative process[9], once the parameters are set the result depends on convergence criteria which should not change throughout the analysis. These criteria depend on the consideration of network complexity and processing resources thus could be regarded as constant per model. Therefore the whole trip distribution procedure is a deterministic – iterative – process for each given set of function parameters.

Traffic assignment is either a multi-step process where the number of steps are previously set (one step for all-or-nothing assignment, and several steps for incremental assignment) or includes an iterative process as well, based on these

fixed steps (equilibrium assignment). The result of a multi-step assignment are determined by network attributes, thus may be considered deterministic. Iterative assignments are also governed by convergence criteria, as the process of trip distribution, and may also be considered a deterministic procedure for the same reasons.

4.2. *Uncertainties*

In several cases of transportation modelling scenarios, estimations are required.

The most relevant example is future travel demand and activity. Common modelling practice dictates the worst and the best case scenarios to assess these uncertainties, however the range of results between those scenarios or the probability of each scenario cannot be readily estimated.

Trip generation determines the expected traffic load of a scenario. The total traffic volume, the number of generated trips is the function of actual data and estimated factors. Its value consists uncertainties, e.g. how many home-to-work trips would manifest given a spread of GDP in a future scenario. Here, either the future value of the GDP or both GDP and the function parameters determining the number of generated trips, may be indefinite.

Although the process of trip distribution is deterministic, the attractiveness, or deterrence of destinations also depend on economic factors, introducing time-variance – therefore liability – to the parameters of the gravity model's deterrence function, thus into the phase of trip distribution.

In mode-choice models there is the possibility that a traveller does not choose the best, most feasible transportation mode for his trip, because personal preferences – e.g. the choice of a scenic route instead of a shorter tunnel – due to the lack of information or other unknown details. The variability of perceived travel costs is also an issue in this process. Stochastic factors influencing modal split are handled by currently available models (Kirchoff, Box-Cox, Logit, etc.) based on extensive survey data.

The primary focus of this paper is to address stochastic effects influencing trip generation. The extension of the process of trip distribution may be achieved in the same method.

It should be noted that although mode choice and choice variance was mentioned here, the experimental models only included a single travel mode – private transport.

5. Synthesis

The inclusion of random variables needs additional considerations after the assessment of deterministic processes and uncertainties. In the current research trip generation potentials were described as the product of several factors, which all had some degree of randomness. Although these factors were determined theoretic-

tically, they may be calculated from real census data. Adaptation of the process is also possible to a different set of variables.

5.1. Structure

Uncertainties at the level of trip generation were implemented both on general and individual levels as follows (Zone and variable structure are shown on Figure 1).

First, production and attraction potentials were assigned for each zone. These attributes represent the basis on which the proportion of generated and attracted trips will be calculated (e.g. the number of households for a residential zone, and the number of workplaces for an industrial, or commercial zone). Values are expressed in ‘trip generation units’.

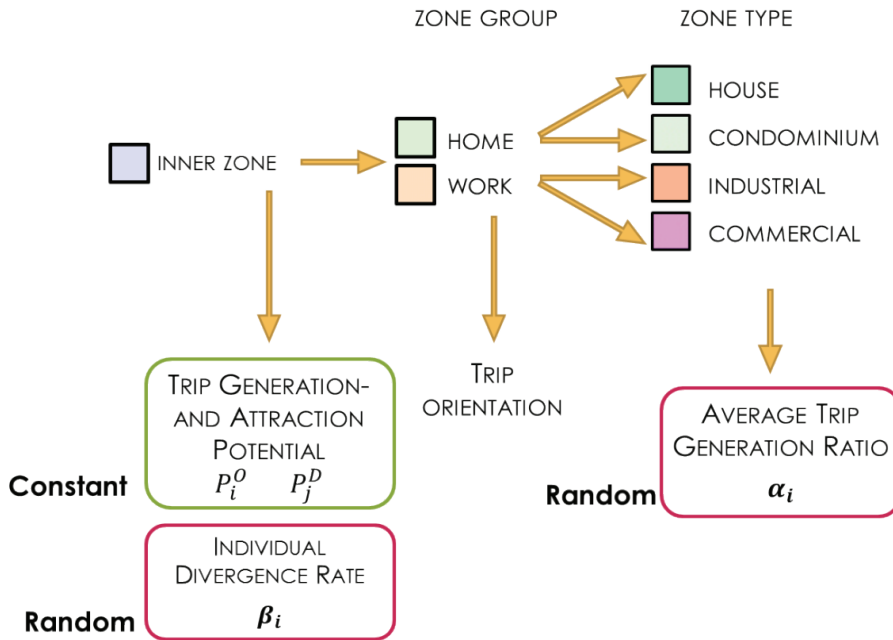


Figure 1. Zone and variable structure

Source: own

Zones are sorted into *zone groups* e.g. ‘home’ and ‘workplace’. These groups both help to assign trip orientations based on the time-of-day and serve as the primary structural level of categorization. Trip orientation was bipolar: ‘home’ zones producing, and ‘work’ zones attracting the trips for a supposed morning period, and having an opposite layout for the evening period. As zone group is a hierarchical and filtering concept, there is no value assigned to this attribute.

Zones are further divided into *zone types*, which – presumably or statistically – generate different number of trips per potential trip generation unit, as – for

example – the average number of trips per household for a suburban area is different than for a range of condominiums. Each zone type have a specific ‘average trip generation ratio’ (trips/trip generation units) to describe this property.

Zone types represent average, general statistical rates of trip generation, but in real life, these values always have a deviance to the average. To account for this discrepancy an ‘*individual divergence rate*’ was introduced on a per-zone basis. This attribute may also help to describe existing differences between similar zones, as one residential area might be more popular than the average, and another be abandoned. The divergence rate is expressed as a percentage.

Application of this structure to the test network is displayed on Figure 2. The test network consisted of 50 links and 25 zones.

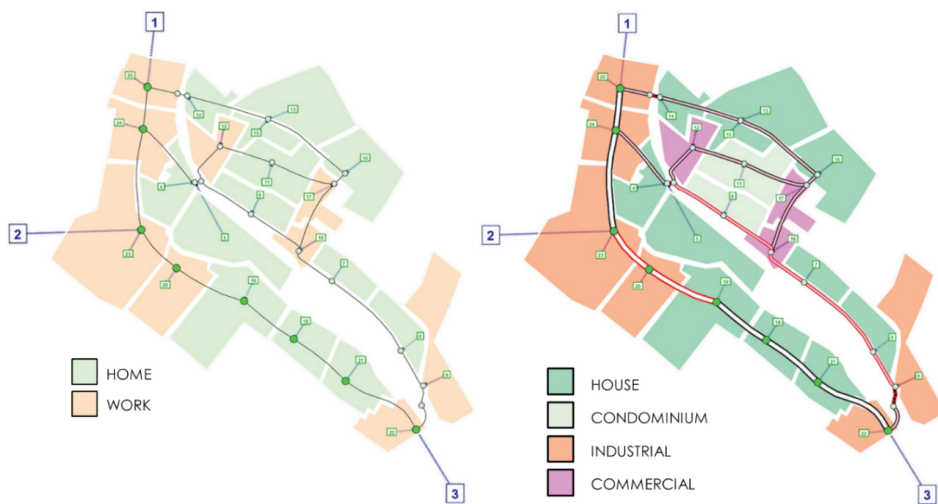


Figure 2. Inner zones by group and type
Source: own

5.2. Random variables

Potentials (number of households, number of jobs) are determinate, and could be derived directly from land use data and population statistics. Average trip generation ratios can be determined by statistical analysis of a time- and/or space varying data set, where the uncertainty – spread, or deviation – of this attribute can also be described. These calculations are outside of the scope of this research, therefore excluded from this discussion. Individual divergence rate expresses either or both the confidence in these average values, and the difference of respective zones.

Each one of these attributes (average trip generation ratio, individual divergence rate) is defined in the model as a random variable, having a probability distribution with describing variables – e.g. a mean and deviation for a normal distribution. Currently, only a simple uniform distribution was set for all variables

(Figure 3), as the objective was to test the suitability of the Monte Carlo method for trip generation.

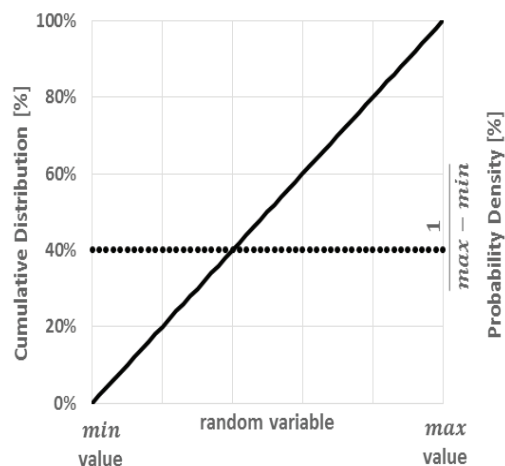


Figure 3. Applied random variable distribution

The range of the average trip generation ratio is presented in Table 1 for each zone type. Note that the 'Outer' zone types are the three zones on the edge of the network model numbered from 1 to 3 inside square frames on Figure 2.

Table 1. Average trip generation ratio per zone type

| <i>Name</i> | <i>Min. value</i> | <i>Max. value</i> |
|-------------|-------------------|-------------------|
| Condominium | 1.1 | 1.5 |
| House | 1.5 | 2.2 |
| Industrial | 0.8 | 0.9 |
| Commercial | 0.9 | 1.0 |
| Outer | 0.9 | 1.1 |

Source: own

For each inner (i.e. non-outer) zone, an individual divergence rate was set. Although the implementation allowed for separate interval definition, this variable had only one range (Table 2), but the respective values were determined separately for each zone.

Table 2. Individual divergence rate of inner zones

| <i>Name</i> | <i>Min. value</i> | <i>Max. value</i> |
|-------------|-------------------|-------------------|
| Inner | 0.9 | 1.1 |

Source: own

The actual value of the attributes are determined before each trip generation procedure by a random number generator, according to these describing variables.

The results of the trip generation - origin and destination traffic volumes – are the product of the introduced random variables and potentials. Average trip generation ratios are presented (1)(2) in a vector format for clarity, but it must be noted that the variance of the individual values is low, vectors consisting only a few zone types' ratios.

$$O_i = \alpha_i \cdot \beta_i \cdot P_i^O \quad (1)$$

$$D_j = \alpha_j \cdot \beta_j \cdot P_j^D \quad (2)$$

where:

- α_i = average trip generation ratio of zone i (thus its zone type's trip gen.r.),
- β_i = individual divergence ratio, zone i ,
- P_i^O = trip generation potential, zone i ,
- P_j^D = trip attraction potential, zone j .

Resulting vectors O and D not necessarily meet the criterion that there are no vehicles lost in the network, anyone who started a trip, finishes it (3).

$$\sum_i O_i = \sum_j D_j \quad (3)$$

Therefore adjustment (4) is required, and was achieved by scaling the firsthand vector D :

$$\acute{D}_j = D_j \cdot \frac{\sum_i O_i}{\sum_j D_j} \quad (4)$$

The origin and destination vectors are then subject to the trip distribution process which currently regarded as a deterministic procedure.

Traffic assignment – as a deterministic process – was not affected by stochastic variables.

6. Implementation

6.1. Simulation cycles

The application of the Monte Carlo method on the transportation network model consists of repetitive trip generation, trip distribution and traffic assignment cycles, fed by different random results of the trip generation phase detailed above. A complete set of cycles is further on referred to as simulation. At the start of each cycle, the values of vectors α and β are randomized to produce a new set of variables.

With Monte Carlo being an unexplored method in the domain of transportation modelling, there is no recommendation about necessary cycle number. Expe-

periments were run with the setting of 100, 1 000, and 10 000 cycles to compare data output resolution and required processing time.

6.2. Platform

PTV's traffic network analysis suite, VISUM provided the modelling environment. This software – like any other modelling application commercially available – was not prepared to process and analyze thousands of successive traffic assignments. A custom utility had to be developed in order to generate the required random variables, and to provide automated network analysis and evaluation. Communication with VISUM was established through the Component Object Model (COM) interface, enabling manipulation of model attributes, reading and saving of separate assignment results.

Generated random values play a major role in Monte Carlo methods, thus for reference it should be noted that for this study the pseudo-random number generator utilized was the default *Rnd()* function of Microsoft's .NET framework (version 4.5).

The description of random variables were implanted in the model as user-defined zone attributes. Origin and destination traffic volumes were managed by the developed software, generating the necessary random variables according to these describing attributes, then performing the necessary arithmetic operations (1)(2)(4) to calculate vectors *O* and *D*.

Execution of the trip distribution and traffic assignment procedures was also handled by the utility, along with the collection of traffic volumes of each link on the network.

6.3. Evaluation

After all cycles of trip generation, trip distribution and traffic assignment had been finished, the results were saved into a macro-enabled excel spreadsheet, for further adjustment. Cumulative distributions and probability densities of traffic volume measurements were determined for each link, using 10 vehicle/hour bins to aggregate the results. Statistical average, standard deviation and variance was also calculated for each link's data set.

7. Results

7.1. Cycle number

Since simulation cycle number have a major effect on both the results' individual values and the processing time, it was logical to assess the effect of cycle number on the results. This effect is demonstrated on a single link's traffic volume results with different cycle numbers (Figure 4, Figure 5, Figure 6).

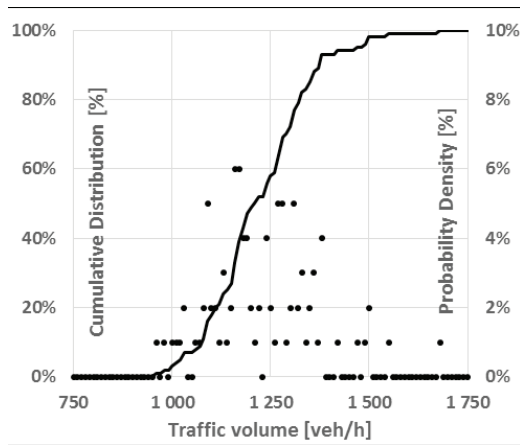


Figure 4. Single link volume results of 100 cycles
Source: own

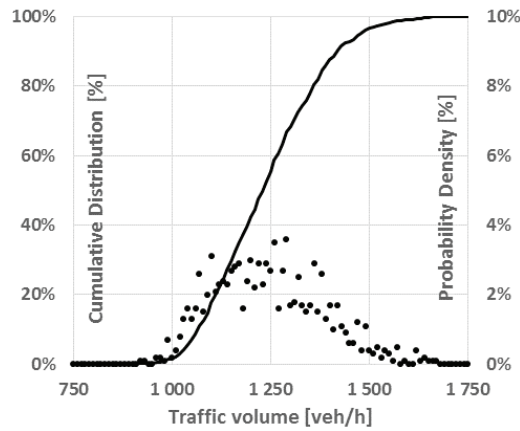


Figure 5. Single link volume results of 1,000 cycles
Source: own

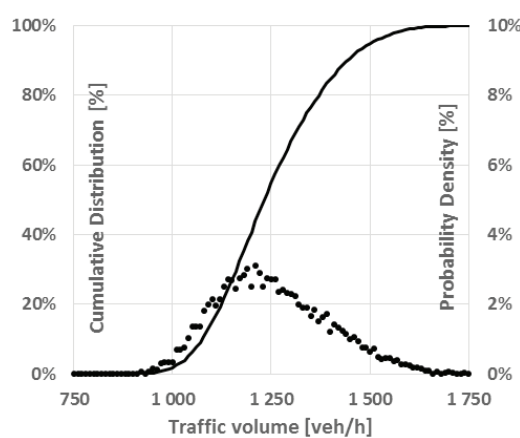


Figure 6. Single link volume results of 10,000 cycles
Source: own

7.2. Link volumes

Note: demonstration of the results is limited to 10,000 simulation cycles, as it produced the most comprehensible data. All links' distributions and densities are demonstrated on Figure 7 and Figure 8 for easy comparison.

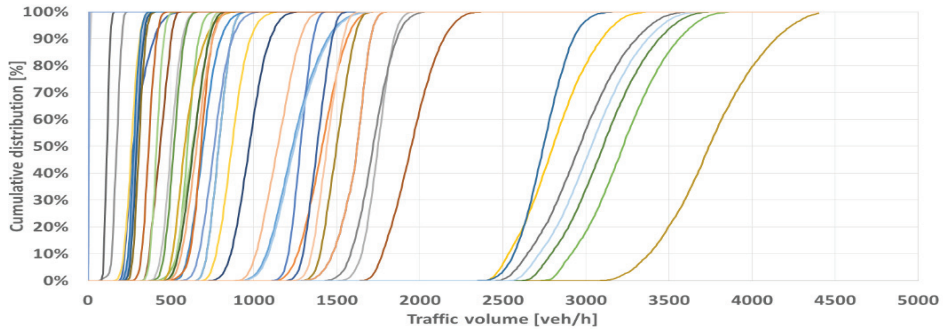


Figure 7. Cumulative distribution of link volumes on the test network

Source: own

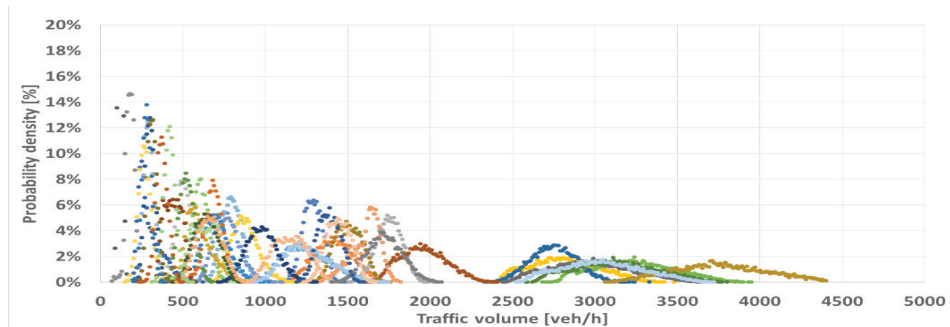


Figure 8. Probability densities of link volumes on the test network

Source: own

Highlighted link volumes are presented here, to be discussed later (see: 8. Analysis).

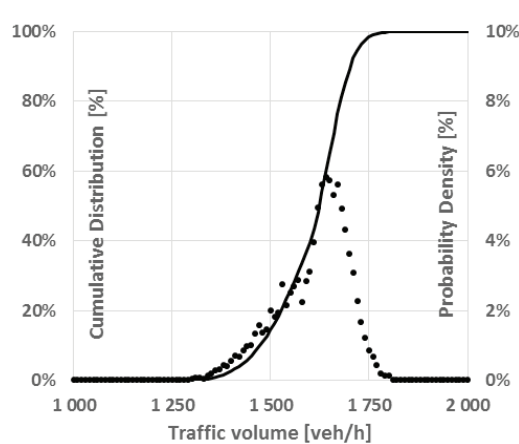


Figure 9. Single link volume results of 10,000 cycles, shape of a skewed Gaussian distribution

Source: own

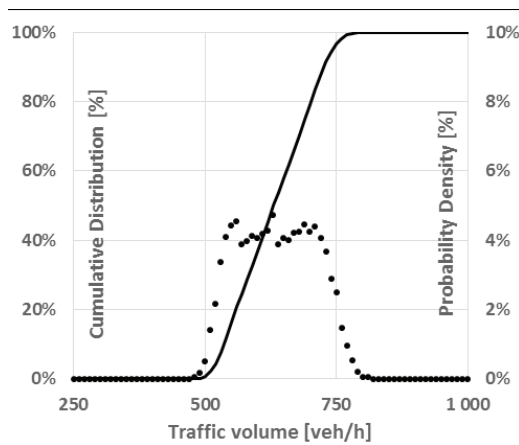


Figure 10. Single link volume results of 10,000 cycles, shape of a rectangular distribution
Source: own

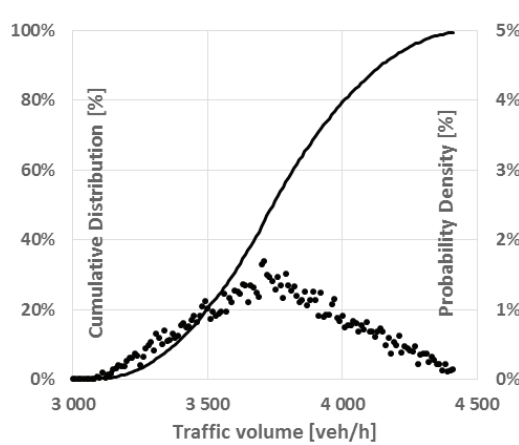


Figure 11. Single link volume results of 10,000 cycles, shape of a triangular distribution
Source: own

The displayed smooth graphs are the results of a 10 000 step cycle, which took approximately 1.5 hours to finish. Final evaluation was completed in one minute. This timeframe expands rapidly by the size of the network. The method described in this paper was later tested on an average-sized city network (7000 links, 74 zones) the same number of cycles taking 70 hours and evaluation about 2.5 hours processing time on the same computer.

8. Analysis

8.1. Cycle number

Inspecting the same link's vehicle throughput results with various simulation cycle numbers, the differences are apparent. By setting a relatively low cycle num-

ber, the shape of a Gaussian distribution is formed through a jagged S-curve of the cumulative distribution graph. Density values are fragmented, which is a clear consequence of the cycle number: the density analysis of $n=100$ measurements cannot be more refined than $1/n$ (see Figure 4).

Increasing the cycle limit's order of magnitude ($n=1,000$) has significant effect on density values, their maximum resolution being 0.1% now. Although the distribution curve is smoother, its overall shape did not change significantly (Figure 5).

Going further by another order of magnitude ($n=10,000$), the density graph also becomes similar to a Gaussian distribution's curve (Figure 6).

Although differences are visual, common statistical values do not show significant contrast in the three data sets (see Table 3).

Table 3. Statistical averages, deviations and variances of the data set in function of cycle number

| n | \bar{v} | σ_v | v_v |
|--------|-----------|------------|-------|
| 100 | 1,223 | 130 | 10.6% |
| 1,000 | 1,238 | 137 | 11.1% |
| 10,000 | 1,247 | 141 | 11.3% |

Source: own

The insignificant increase of the average is deceptive. Its value varies between simulations due to the stochastic nature of input variables, hence inclination is only apparent.

8.2. Link volumes

Different links on the network show different shapes of cumulative distribution and probability density graphs, meaning different reactions to traffic input. Since traffic assignment also includes shortest – or cheapest – path search between each zone pair, these graphs also indicate excessive volume-capacity ratios and critical roles of respective links.

Graphs on Figure 6 are related to a link connecting the inner road network to the bypass. This link has a nearby alternative with same features (capacity, speed restriction) but this link provides the higher accessibility of the two, and also channels higher traffic demand. Curve shape is close to a Gaussian distribution's.

A skewed bell-shape on Figure 9 belongs to a link of a similar connecting role, but without nearby alternatives – actually one of its distant alternatives is the link mentioned before.

A link with low traffic demand shows the shape of a rectangular distribution on Figure 10. This link is the neighboring alternative of the first example (Figure 6), with lower importance (providing accessibility to fewer zones) on the network.

Links shaping a triangular distribution, with highly spread traffic volume range are links with low volume-capacity ratios (hence low saturation), as one of the bypass sections shown on Figure 11.

8.3. Overview

To have an overview of the whole network, all link distribution curves and density graphs are displayed on Figure 7 and Figure 8. These figures show a wide range of link volumes, higher capacity – and higher volume – links having a relatively low slopes, and low capacity links having steep slopes.

It is important to note that although the input variables had a uniform – i.e. rectangular – distribution, most volume graphs bear some resemblance to a normal distribution. Speculations lead to the Central Limit Theorem, which states that the average of a large enough set of random independent trials lead to a normally distributed result. This theory is best demonstrated by a Galton board. The relevancy of this example is in the link volume's composition regarding paths connecting different zone pairs. In this context a refined bell shape means that the corresponding link volume consists of routes from a large number of zones, as those volumes were determined by the product of two uniformly distributed random variables (see equations (1) and (2)).

9. Conclusions

Results of regular transportation network model analysis consist of several variables per network element. Each of these having a singular value e.g. a vehicle volume on a road section. By introducing a stochastic element at the level of trip generation in a hypothetical test network – thus affecting the total and route specific vehicle volume of the network in a controlled random fashion – it was possible to gain more detailed information about link volumes than ordinary analysis would allow – including the best and the worst case scenarios. The calculated expected – average – volume and its deviation describe the range of possible outcomes, and could explain certainty of the results instead of a singular value. The increased number of simulation cycles add fine details to this description at the cost of higher calculation time and careful evaluation of the results.

The ongoing research uncovered a new aspect: At the level of link volumes, range and distribution show high diversity. The total range of traffic volume varies in width, and probability densities might take the shape of rectangular, triangular, Gaussian, and skewed Gaussian distributions, which show the different exposition of different links to the effect of stochastic input variables denoting the emergence of a higher level context: network vulnerability. Further research steps are required to analyse this perspective.

The demonstrated procedure is easily parallelized, by distributing the cycles to several computers, running the process simultaneously. Results could be evaluated after concatenating the output of these individual simulations.

This novel approach to transportation network analysis promises the inclusion of certainty in the results of traditional models by giving a probability distribution to the affected outputs. Application of the methodology to a detailed and properly calibrated model would push its already reliable results to a higher level.

10. Future aspects

The current focus was set on the process of trip generation. Involving trip distribution in the methodology is a logical path. A widely used example is the gravity model, where a deterrence function describes the attractiveness of destination zone j to persons entering the network from zone i :

$$T_{ij} = O_i \cdot D_j \cdot f(C_{ij}) \quad (5)$$

where:

- $f(C_{ij})$ = deterrence function,
- C_{ij} = travel cost between zones i and j .

An implementation of the deterrence function by the familiar combined-logit-model is:

$$f(C_{ij}) = a \cdot C_{ij}^b \cdot e^{-c \cdot C_{ij}} \quad (6)$$

where:

- a, b, c = calibration factors.

Variables of the deterrence function – on which the distribution of the previously generated trips relies – are also estimations. Here, the calibration factors a, b, c (see (6)) may also be described as random variables to express the uncertainty of the traveller's destination choice.

Modal-split models are already established, the higher complexity solutions account for the stochastic nature of traveller's mode choices. These models may not be a suitable environment to introduce MC principles, but it may worth a trial to conclude that path. These models have similar calibration factors to the gravity model's deterrence function in trip generation (see (6)) which means that implementation would not be more difficult.

Traffic assignment (all-or-nothing, incremental, equilibrium) has no clear opening for adaptation. All-or-nothing and incremental assignments are essentially single or repetitive shortest path searches, having no relevancy to real-life deci-

sions as the shortest path searches do not result in multiple alternative routes. Only the shortest routes are determined therefore it is not possible to include the stochastic nature of user choice. It would also be almost impossible to manipulate the shortest path search algorithms of existing transportation modelling software, being highly optimized, and non-customizable and also inaccessible part of the software.

Equilibrium assignment algorithms use several variables to check convergence – none of them could be interpreted as the model's uncertainty. Also there are stochastic equilibrium assignment methods available to model the probabilistic nature of user choice, which would make the MC implementation obsolete.

As it was mentioned before, description of network vulnerability may be possible by making use of the MC simulation results. This aspect definitely needs additional extensive research taking network topology into account.

References

- [1] Rohács D., Non-Linear Prediction Model for the European Small Aircraft Accessibility for 2020. (PhD dissertation), Budapest: BME, 2007.
- [2] Anderson H. L., Metropolis, Monte Carlo, and the MANIAC. Los Alamos Science, pp. 96-107, 1986.
- [3] Zio E., The Monte Carlo Simulation Method for System Reliability and Risk Analysis. London: Springer-Verlag, 2013.
- [4] Dimov I. T., Monte Carlo Methods for Applied Scientists. Singapore: World Scientific, 2008.
- [5] Dias M. A. G., Quasi-Monte Carlo Simulation. [Online]. Available: http://marcoagd.usuarios.rdc.puc-rio.br/quasi_mc.html. [Accessed 02 12 2013].
- [6] PTV AG, PTV VISUM 13.0 Fundamentals, Karlsruhe: PTV AG, 2013.
- [7] Bell M. G. H., Iida Y., Transportation Network Analysis. Chichester: Wiley, 1997.
- [8] Ortúzar J. d. D., Willumsen L. G., Modelling Transport. John Wiley & Sons, Ltd., 2011.
- [9] Schnabel W., Lohse D., Grundlagen der Straßenverkehrstechnik und der Verkehrsplanung. Berlin: Kirschbaum - Beuth, 1997.