BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
FACULTY OF CIVIL ENGINEERING
DEPARTMENT OF HIGHWAY AND RAILWAY ENGINEERING

ANALYSIS AND QUANTIFICATION UNCERTAINTY OF THE TRAFFIC ASSIGNMENT

Ph.D. thesis booklet

MUNDHER ALI SEGER

Supervisor:

LAJOS KISGYÖRGY, PhD

Associate Professor

Budapest, 2020

1. Background and Motivation

Transportation planning aims to forecast people's travel behaviours to supply adequate facilities and services to satisfy future demand. Travel Demand Models (TDM) have been developed in the 1950s for future travel demand forecasting and used as decision support tools by the transport planners over the last several decades. TDM, like any other model in applied science and engineering, cannot describe real systems precisely because the information of the underlying model is incomplete and insufficient or due to the stochastic variability of the inputs data and the model itself, which leads to uncertainty in executing the model. Such causes uncertainty on both the model parameters' values and on the hypotheses forming the model structure. Consequently, uncertainty propagates through the TDM and produces variability in its outputs.

Thus, uncertainty is unavoidable in the modelling TDM, which is very important when attempting to understand travel behaviour. The importance of uncertainty is in revealing confidence in a TDM and its predictions. Quantitative uncertainty analysis promotes the separation of assessment and decision-making. Enabling decision-makers to evaluate the full range of possible outcomes turns interpretation of predictions into a matter of policy. Hence, uncertainty estimates concisely convey information to a decision-maker about the degree of confidence that can be placed in any point estimates. The importance of uncertainty is in revealing confidence in a model and its predictions. Quantitative uncertainty analysis promotes the separation of assessment and decision-making. Enabling decision-makers to evaluate the full range of possible outcomes turns interpretation of predictions into a matter of policy. Consequently, Confident predictions of future travel behaviour can only be made by taking uncertainty into account. The inherent uncertainty in TDM, if not adequately quantified, makes analysis highly unreliable.

The source of uncertainty can be categorised into two types according to the possibility of handling. The first type is the aleatoric uncertainty (also called: stochastic uncertainty, irreducible uncertainty, or variability), it occurs due to inherent randomness or variety, and it can happen among divisions of a population or due to spatial or temporal variations. Therefore, this type of uncertainty cannot be eliminated or reduced or controlled. The second type is the epistemic uncertainty, which is come from a lack of knowledge, incomplete information, or ignorance in any phase of operation of a design process. This type of uncertainty's primary cause is incomplete or insufficient information on some aspects of the model. Therefore, an increase in knowledge or information can lead to a decrease in the epistemic uncertainty. In TDM, both aleatory uncertainty and epistemic uncertainty are existing simultaneously. The significant problem of uncertainty in TDM arises from many input variables that mainly come from the OD matrix. As a result, it is necessary to account for all epistemic uncertainties caused by the OD matrix to obtain more realistic traffic flow forecasts in the future.

Uncertainty can be found at all stages of the TDM stages, where the errors passed from stage to others and propagated over the whole model. Therefore, studying the uncertainty in the last stage, which is represented by the traffic assignment model, is more important because it draws the result of uncertainty in the whole TDM. The traffic assignment model aims to distribute the OD flows throughout links of a given transport network, and according to methods, depends on either specific deterministic or stochastic assumptions. Moreover, to simulate the travel demand interactions with transport system supply for producing an output that describes the transport system's mean state. Therefore, the uncertainty of the traffic assignment model is reflecting the uncertainty of the OD flows.

The outputs of the traffic assignment model are generally significant to decisions related to transport management and strategy. So, the simple point estimates, even if they refer to the most probable values, are not sufficient and satisfactory for a precise and safe assessment of the risks associated with such decisions. Therefore, the need for quantifying the uncertainty which is related to traffic assignment model estimates is strongly present. Although there are different studies had addressed the topic of uncertainty in the traffic assignment model and note that a limited number of studies consider the quantification of uncertainty in the form of confidence intervals or other similar measures. On the other side, most of these studies took into consideration the model-uncertainty rather than the input-uncertainty. Besides, the studies that dealt with input-uncertainty were investigated mainly in two input variables; time and cost With well-known previous studies in mind, which is show a lack of knowledge in analysing input-uncertainty for the OD matrix.

Given the above facts, there is a significant necessity for more robust estimate outputs of the traffic assignment model by adding confidence intervals to these point estimates. The estimates with confidence intervals will enhance the traffic assignment model's ability to describe the stochastic variability of travel behaviour more precisely. Such estimates with confidence intervals required full information about the statistical distributions for all input variables represented by the OD matrix, to produce confidence intervals for traffic assignment outputs.

Based on the previous, it becomes clear the importance of developing a practical framework for analysing and quantifying uncertainty in the traffic assignment model caused by the input OD matrix. This framework will assist transportation modellers, planners, and decision-makers, to have a fresh look at the uncertainty in traffic assignments of transportation models. The proposed framework consists of two approaches. The uncertainty analysis approach explores and evaluates the confidence in the traffic assignment model and its predictions. On the other side, the uncertainty quantification approach aims to determine the confidence intervals for the traffic assignment outputs.

2. Hypothesis, Research Questions and Study Objective

After the background and motivation in the previous section, which describes the general study problem. I can introduce my hypothesis, research questions and objective of the work.

Hypothesis:

The OD matrix is not sufficient for entire representing the travel demand model because of the outputs of the traffic assignment model involving different types of uncertainty due to the stochastic variability of the travel behaviour between origin and destination zones.
Therefore, the travel demand model required additional input data for representing the variability of travel behaviour to quantify the uncertainty of the traffic assignment model.

This hypothesis focuses on the possibilities of adding the variances of OD pairs in the traffic assignment model for analysing and quantifying the traffic flows' uncertainty. The challenges here are: define the necessary information that useful in analysing and quantifying the uncertainty of traffic assignment model, and how to find out the distribution parameters (mean and variance) of each OD pair.

Research Questions:

Based on the hypothesis, four research questions appear. The answering of these four questions will provide the required information for the study objective. The four research questions of this work are:

1. How the stochastic variation of the OD matrix affects the traffic assignment model outputs?
2. What is the set of OD pairs that generate traffic flow in each link on the transport network?
3. What are the choice proportions over each link on the transport network by the OD pairs?
4. What are the distribution parameters of each OD pair within the OD matrix?

Study Objective:

The general objective of this thesis is to develop a methodological and practical framework for analysing and quantifying traffic flows uncertainty in traffic assignment model.

3. New scientific results

Thesis I:

I have developed a methodology capable of analysing and classifying the uncertainty in the traffic assignment caused by the stochastic variation of the OD matrix by using the Monte Carlo Simulation method.

The developed methodology was built using the Monte Carlo simulation method to generate multiple OD matrices and the traffic assignment model via VISUM software to produce traffic flows for whole links on the transportation network. The structure of this methodology consists of six stages:

1) Input stage: The required data for this stage represented by the base OD matrix (OD^{base}) and the observed traffic flows (q^{obs}) on links.

2) MC simulation process stage: In this stage, the simulation code has been written using both MATLAB software and VISUM software in purpose to predict traffic flows on the transport network. The simulation code involves two processes: the first is the generation of the multiple OD matrices by MC method for a given type of probability distribution and standard deviation and based on the base OD matrix.

3) Analysis of predicted traffic flow stage: The results of the simulation process are analysed to find statistical characterisations of traffic flows attributes. The number of obtained attributes is equal to the number of iterations (N) used in the simulation. The statistical characteristics defined by two parameters; the average predicated value results ($\bar{q}_l$), and the variability (σ_l).

4) Predictive uncertainty stage: In any uncertainty analysis process, setting limits for the predictive uncertainty required to increase understanding about the model's behaviour in both bias and variability. The GEH statistic has been used as a limitation of the bias in this study. The GEH statistic is a form of Chi-squared statistic that can be used to compare observed and modelled counts as Equation (1):

$$GEH = \sqrt{\frac{2\left(q_l - \bar{q}_l\right)^2}{\left(q_l + \bar{q}_l\right)}} \tag{1}$$

For this study, GEH statistic bands less than (5) is used to explain the bias limit.

By solving Equation (1), the lower and the upper bias limit as follows:

$$q_l^{low} = \bar{q}_l + 6.25 - 1.25\sqrt{16\,\bar{q}_l + 25} \tag{2}$$

$$q_l^{up} = \bar{q}_l + 6.25 + 1.25\sqrt{16\,\bar{q}_l + 25} \tag{3}$$

Where: q_l^{low} lower allowable bias limit for the link (l),

q_l^{up} upper allowable bias limit for the link (l);

If the observed traffic flows $\left(q_l^{obs}\right)$ lying within the variability (σ_l), then these are precise. Otherwise, these observed traffic flows $\left(q_l^{obs}\right)$ are imprecise.

If $\bar{q}_l - \sigma_l \le q_l^{obs} \le \bar{q}_l + \sigma_l$ Then low variety (precise) (4)

If $\bar{q}_l - \sigma_l > q_l^{obs} < \bar{q}_l + \sigma_l$ Then high variety (imprecise) (5)

In the same way, if the observed traffic flows $\left(q_l^{obs}\right)$ are lying within upper and lower bias limits, then these are accurate. Otherwise, these observed traffic flows $\left(q_l^{obs}\right)$ are biased.

If $q_l^{low} \le q_l^{obs} \le q_l^{up}$ Then low bias (accurate) (6)

If $q_l^{low} > q_l^{obs} < q_l^{up}$ Then high bias (inaccurate) (7)

5) Uncertainty classification stage: The classification has been subdivided into four cases according to bias and variability characteristics of the predictive traffic flow. The proposed classification was consisting of four cases of uncertainty:

- Case I: Accurate and precise (Low uncertainty); this case occurred when the observed traffic flows have low variability and low bias.
- Case II: Accurate and imprecise (Moderate uncertainty); this case occurred when the observed traffic flows have high variability and low bias.
- Case III: Inaccurate and precise (High uncertainty with ensemble agreement); this case occurred when the observed traffic flows have low variability and high bias.
- Case IV: Inaccurate and imprecise (High uncertainty with divergent estimates); this case occurred when the observed traffic flows have high variability and high bias.

6) Uncertainty visualisation stage: The major role of this visualisation of uncertainty is to give information about the level of uncertainty to the decision-makers. Therefore, in the visualisation stage, a specific colour was given for each case of uncertainty. In this methodology, the statistical and probability-based visualisation method is used to visualise uncertainty with a bar chart or probability distributions. In practice, all these four cases can appear with different likelihood proportions. Figure 1 shows an example of a probability density curve for predicted traffic flow for a link (l). The coloured areas under the curve represent the predictive uncertainty cases. Figure 2 shows an example of uncertainty analysis for five links appearing mixed uncertainty cases in each one.

The relationships connecting these stages are presented in Figure 3.

Related publication: [J-1], [C-1]

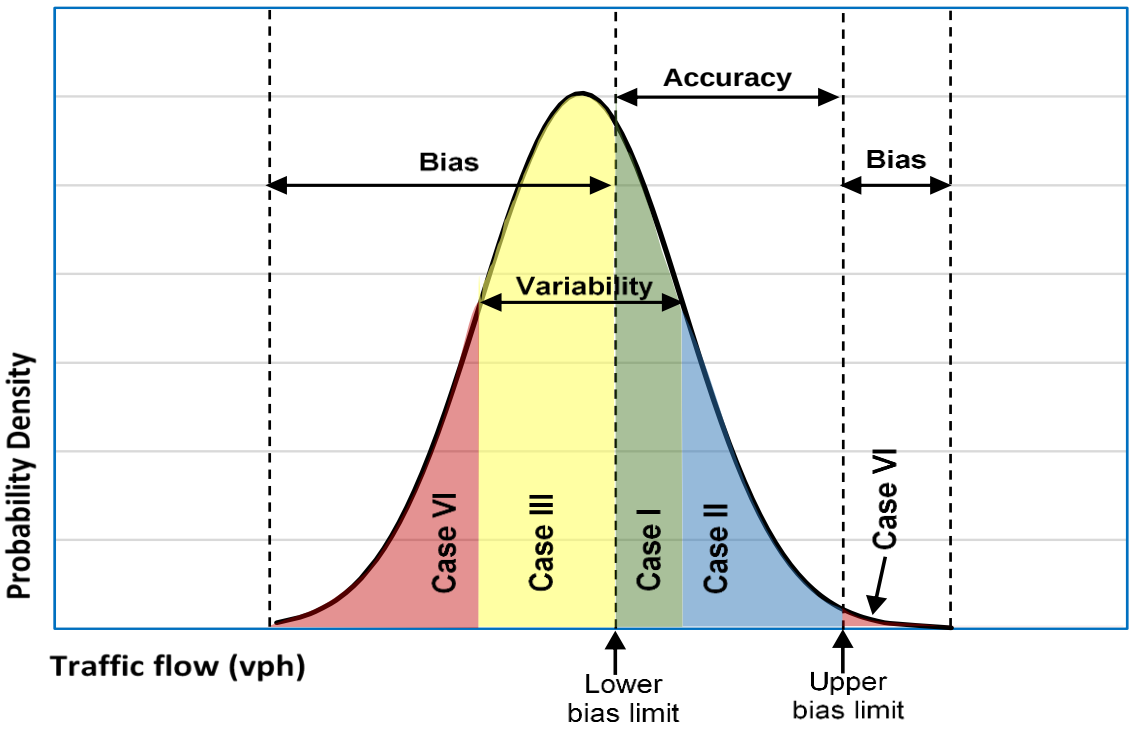

Figure 1: Shows the probability density of observed traffic flows addressed with the uncertainty cases.

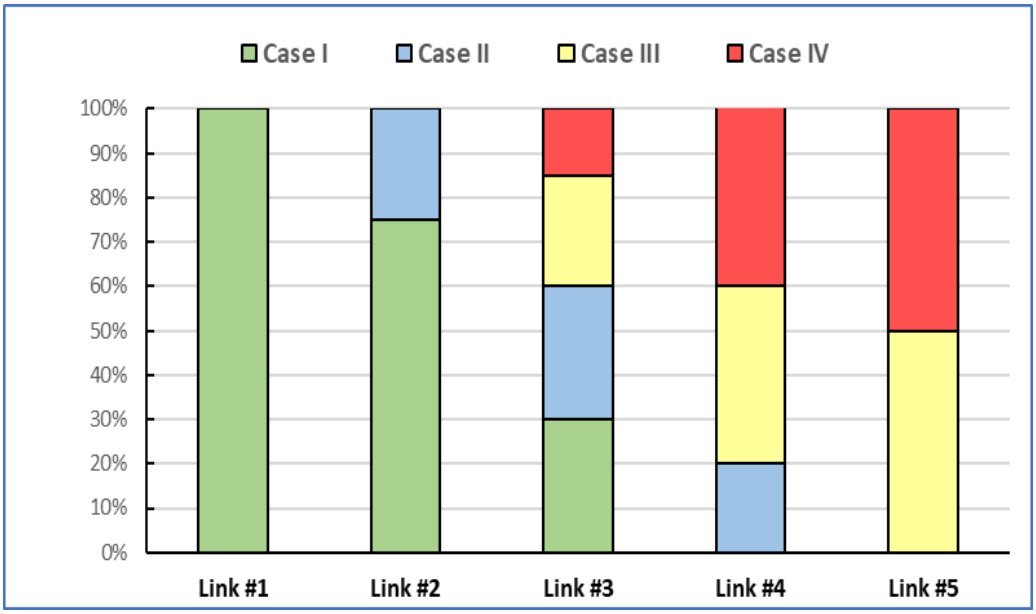

Figure 2: Shows the percentage of the uncertainty cases of the traffic flows for several links.

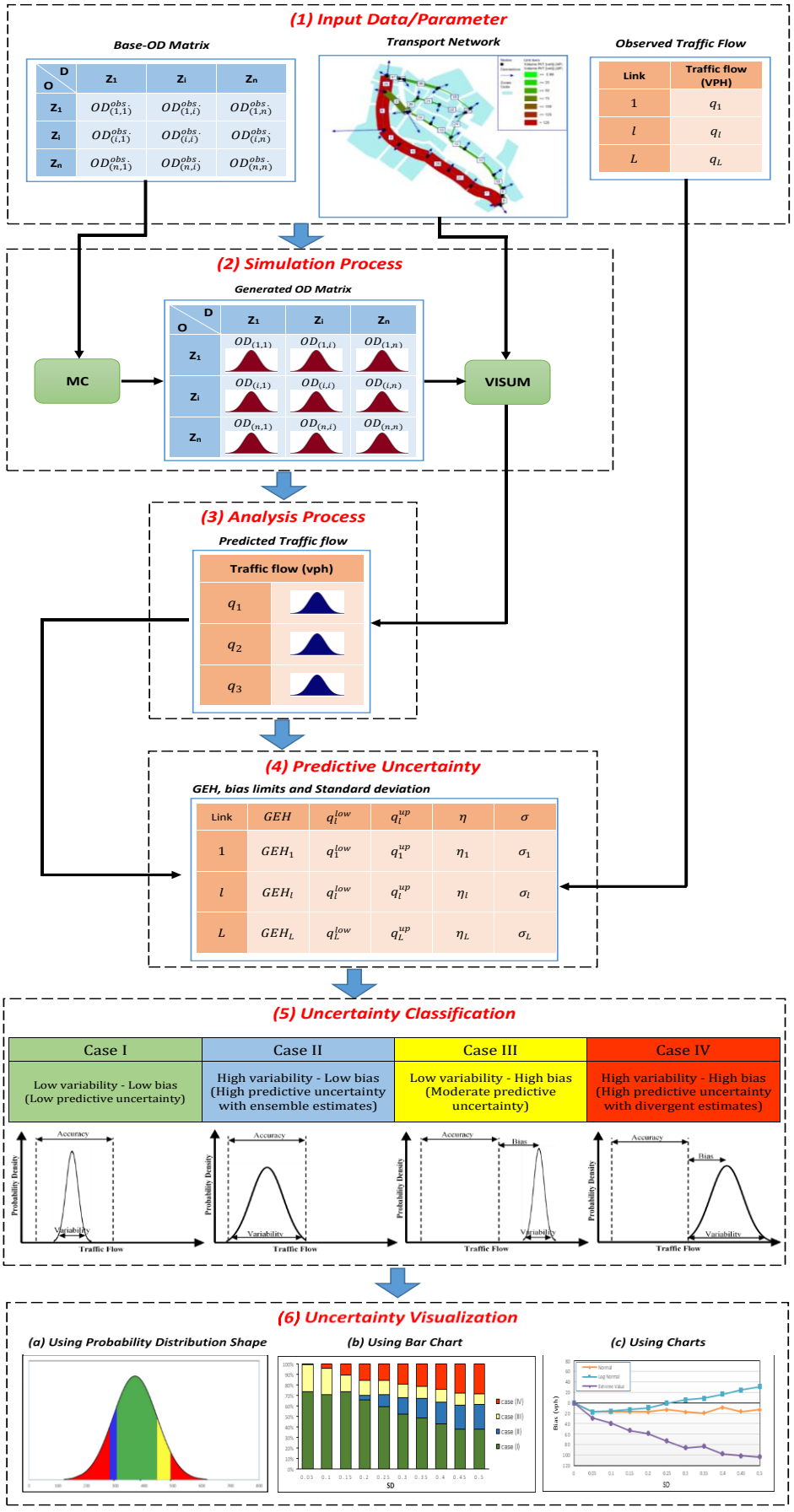

Figure 3: Flowchart of uncertainty analysis in traffic assignment using MC simulation.

Thesis II:

I have innovated a new algorithm (GSA-TAM) designed to analyse the sensitivity of the input data of the traffic assignment model. The (GSA-TAM) algorithm aims to identify the set of OD pairs that creating traffic flow in each link on the transport network and determine the influence scope of each OD pair on traffic flow in every link on the transport network.

The Global Sensitivity Analysis for the Traffic Assignment Model (GSA-TAM) algorithm is built on both the variance-based sensitivity analysis method, one of the global sensitivity analysis methods, and the traffic assignment model via VISUM software. Besides, the simulation process of the (GSA-TAM) algorithm has relied on Quasi-Monte Carlo (QMC) method to generate high-dimensional OD matrices with low-discrepancy sequences. The primary purpose of the GSA-TAM algorithm: First, identifying the set of OD pairs that creating traffic flow in each link of the transport network. Second, determining the influence ratio of each OD pair on traffic flow variance in every link on the transport network.

The GSA-TAM algorithm has been divided into five-stages to be clear and straightforward in implementation for any network. The relationships connecting the methodology stages are presented in Figure 4. The GSA-TAM algorithm stages are as follow:

Definition stage: Three things should be defined for this stage; (i) Number of OD pairs (p) in the travel demand model, (ii) Number of links of the transport network (L). (iii) The sample size of the simulation process (N).

Generation stage: This stage includes building two sets of matrices (M_1) and (M_2), using the QMC method. In both matrices, the number of columns represents the number of OD pairs (p), while the number of rows depicts the sample size (N).

Running stage: In this stage, the GSA-TAM algorithm has been coding by MATLAB software and linked with VISUM software. The results of model running are two sets of vectors (Q_1) and (Q_2), corresponding to the two sets of matrices (M_1) and (M_2) respectively.

Calculation stage: During this stage, Equation (8) and Equation (9) are used to calculate the first-order sensitivity (S_i) indices and total sensitivity indices (ST_i), respectively.

$$S_i = \frac{\left[\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(q_k^{M_1}\right)^2 - f_o^{\,2}\right] - \left[\left(\frac{1}{2N}\right)\sum_{k=1}^{N}\left(q_k^{M2} - q_k^{M_3}\right)^2\right]}{\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(q_k^{M_1}\right) - f_o^{\,2}} \quad (8)$$

$$ST_i = \frac{\left[\left(\frac{1}{2N}\right)\sum_{k=1}^{N}\left(q_k^{M1} - q_k^{M_{3_i}}\right)^2\right]}{\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(q_k^{M_1}\right) - f_o^{\,2}} \quad (9)$$

Where: $q_k^{M_1}$, $q_k^{M_2}$ and $q_k^{M_3}$ are model output for any parameter (i) in the sample (k) corresponding to matrices M_1, M_2 and M_3 respectively, the M_3 is also a matrix formed from all the elements of the matrix M_2, except the i column, where elements are taking from the matrix M_1. The f_0^2 is a constant calculated as follow: $f_0^2 = \left(\frac{1}{N}\sum_{k=1}^{N} q_k^{M_1}\right)^2$.

Analysis stage: In the last stage, the link choice set $(\mathcal{F}_l)$ will be identified. Moreover, determining the scope of effectiveness for each OD pair on the transport network.

Related publication: [J-2], [J-4]

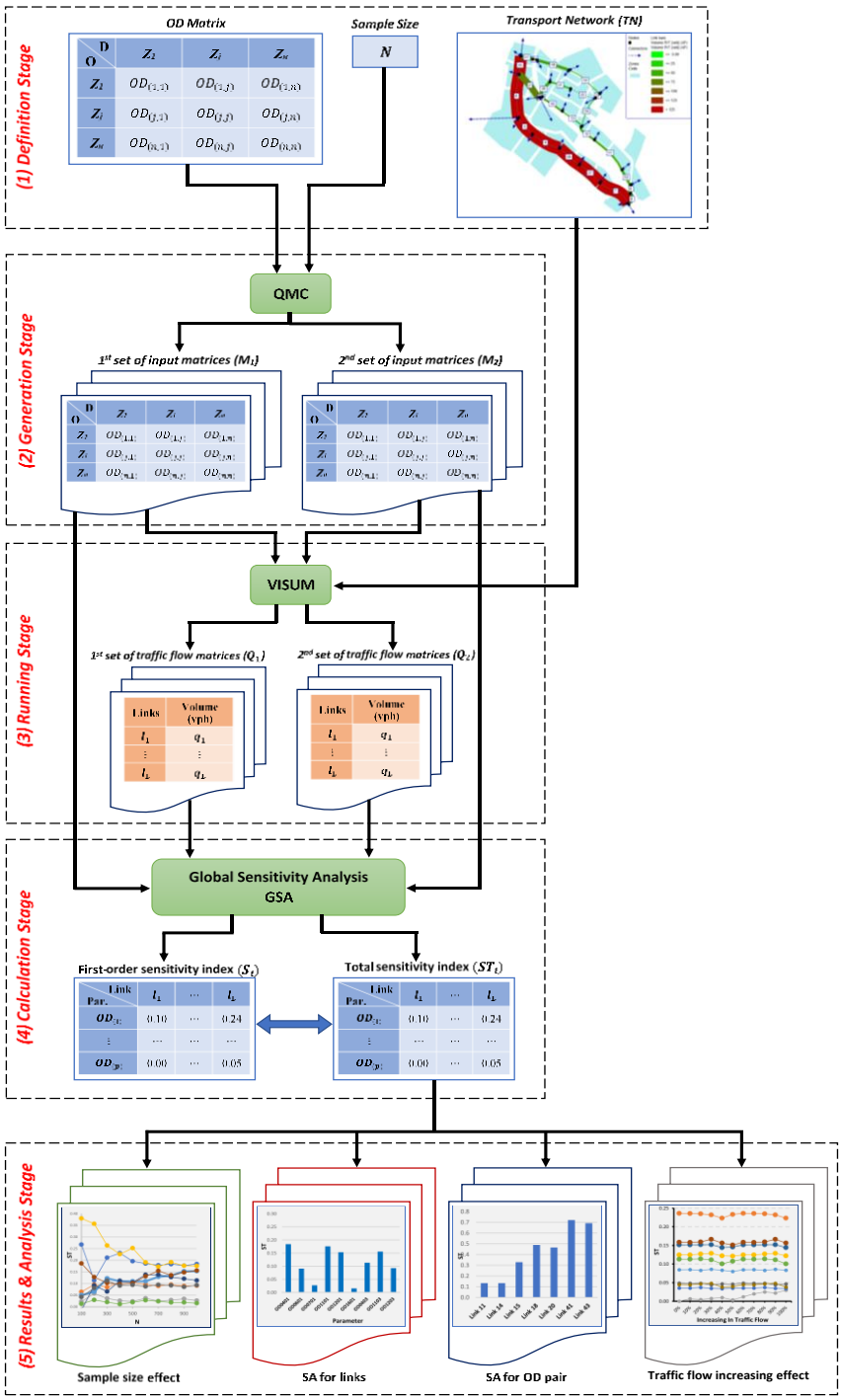

Figure 4: GSA-TAM algorithm

Thesis III:

I have developed a new method for estimating the link choice proportions using the Monte Carlo simulation method and based on the Maximum Likelihood Estimation framework.

The reason for innovating this method is to provide information fill the gap of knowledge about what are the proportions of the OD pairs of the link choice set that using this link (i.e., link choice proportions). This method is created by optimisation of the traffic assignment model in the Maximum Likelihood Estimation MLE framework.

The importance of this method lies in the use of link choice probabilities in uncertainty analysis for the traffic flow in each link separately, unlike other methods that do not provide this technique because they depend on route choice proportions. The new methodology consists of three stages:

1) Input data stage: the required data for this stage consists of three components: the base OD matrix, the sample size for the simulation process (N), and the link choice set ($\mathcal{F}_l$).
2) Generation stage: a MATLAB code has been written and linked with VISUM software. Hence, the simulation code in this stage involves two processes: the first is generating (N) of the OD matrix by the MC method for a given type of probability distribution and standard deviation via base OD matrix (OD^{base}). While the second process is the process of entering the generated OD matrix and the transport network in the VISUM to produce the predicted traffic flow attributes for the transport network.
3) Parameter estimation: In the last stage of this methodology, the link choice proportions will be estimated using Gaussian Process GP based on the Maximum Likelihood ML framework by Equation (9):

$$\widehat{\beta_l} = (\mathcal{F}_l{}^T \mathcal{F}_l)^{-1} \mathcal{F}_l{}^T q_l \qquad (9)$$

Where:

$\widehat{\beta_l}$ Link choice proportions,
$\mathcal{F}_l$ Link choice set,
q_l calculated traffic flows.

The relationships connecting these stages of the methodology are presented in Figure 5.

Related publication: [J-2]

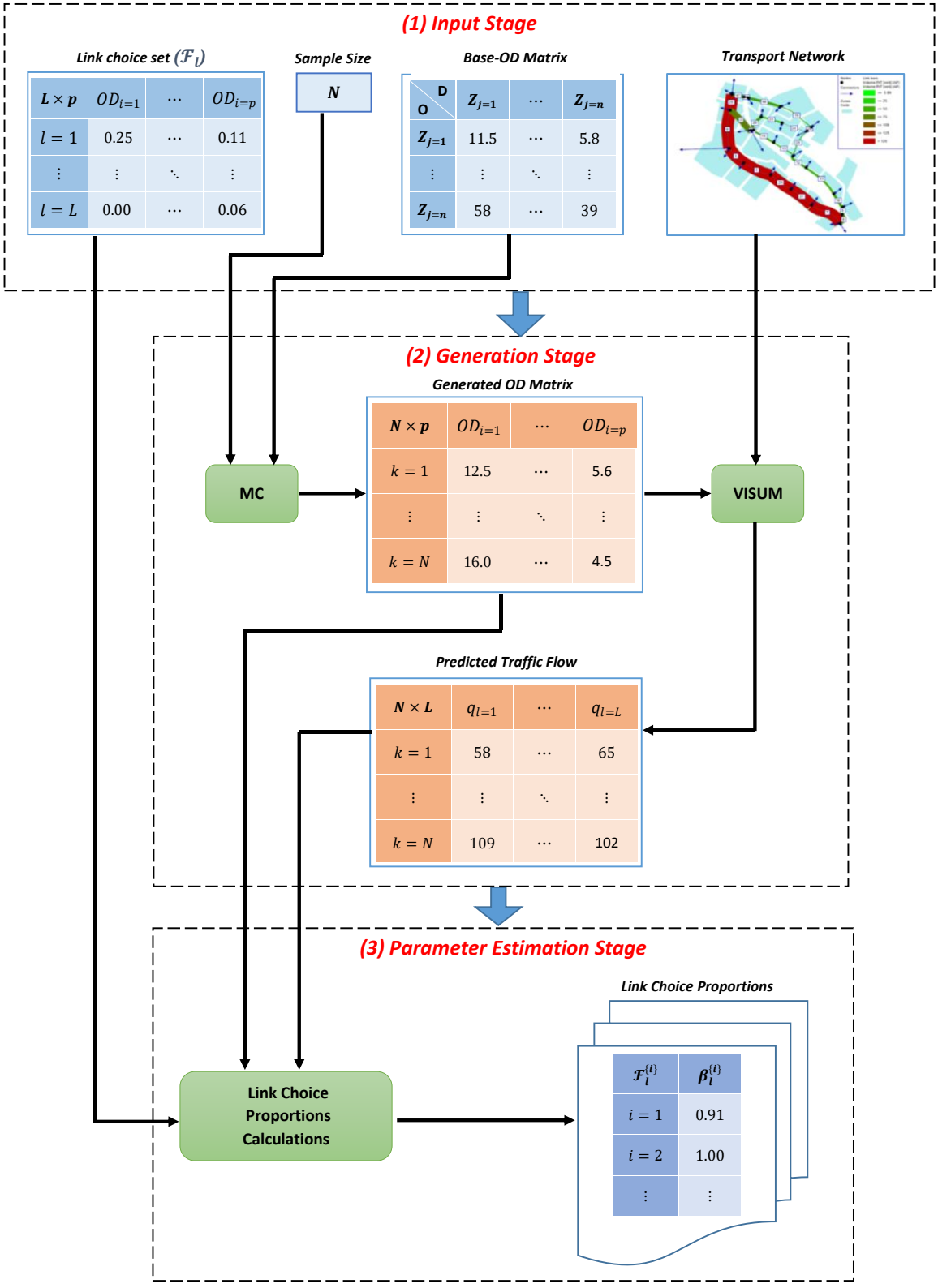

Figure 5: Flowchart illustrates the methodology of estimation of the link choice probabilities.

Thesis IV:

I have innovated a new method to estimate the distribution parameters of the OD matrix from the observed traffic flow data.

The key aim of this method represents in; First, calibrating the mean values of the OD pairs to reproduce the average-OD matrix. Second, estimating the variance values of the OD pairs to create the variance-OD matrix. This method has built by applying the Inverse Uncertainty Quantification (Inverse-UQ) technique to extract the statistical parameters for the input data (i.e., OD matrix) from the real data (i.e., traffic flow) through the traffic assignment model. This method was built on four components: (1) Least Square Adjustment (LSE) method, (2) Traffic assignment model via VISUM, (3) MC simulation method, and (4) traffic flow data. This method consists of three stages:

1) Input stage: the required data for this stage consists of: (i) The observed traffic flow data ($q^{obs.}$). (ii) The base OD matrix (OD^{base}). (iii) The sample size of observations hours (N). (iv) Furthermore, Define a Minimum Residual Sum of squares (Min. RSS) as a benchmark to set the accuracy limit of the results, where the RSS is a measure of the difference between the data and an estimation model.
2) Calibration stage: In this stage, a MATLAB code has been written and linked with the VISUM software. The Inverse-UQ technique based on LSA has been used to calibrate the input base OD matrix (OD^{base}).
3) Calculation stage: The results of the calibration stages are used to establish the average-OD matrix ($OD^{ave.}$) and the variance-OD matrix ($OD^{var.}$). For example, Figure 6 shows the result of the calibration process for the OD pair from zone 06 to zone 01 (OD_{06-01}). The right side shows the probability distribution (OD_{06-01}), and the left side shows the scatter-plots of the calibrated OD pair $\left(OD^{clb.}_{06-01}\right)$ vs the observed traffic flow ($q_2^{obs.}$), the results are surrounded by 68% and 95% confidence intervals boundaries.

The relationships connecting these stages of the methodology are presented in Figure 7.

Related publication: [J-3]

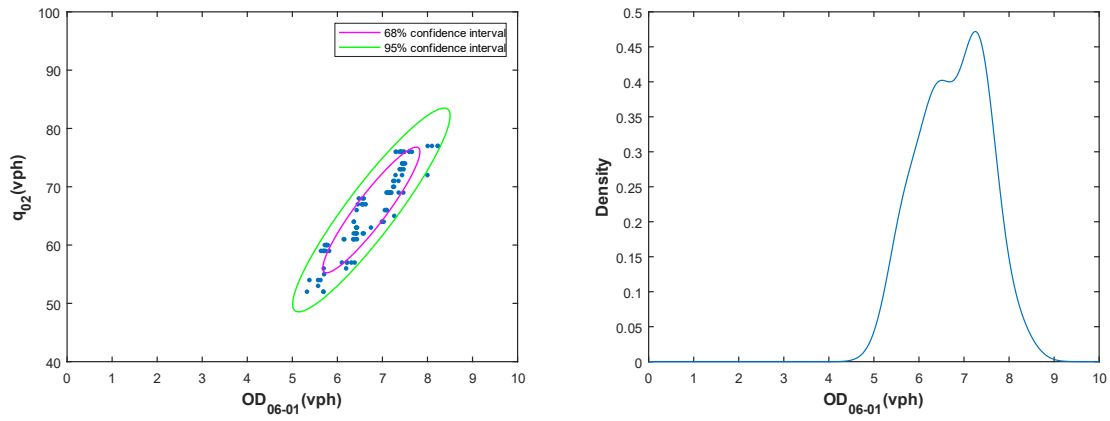

Figure 6: shows the result of the calibration process for the (OD_{06-01}).

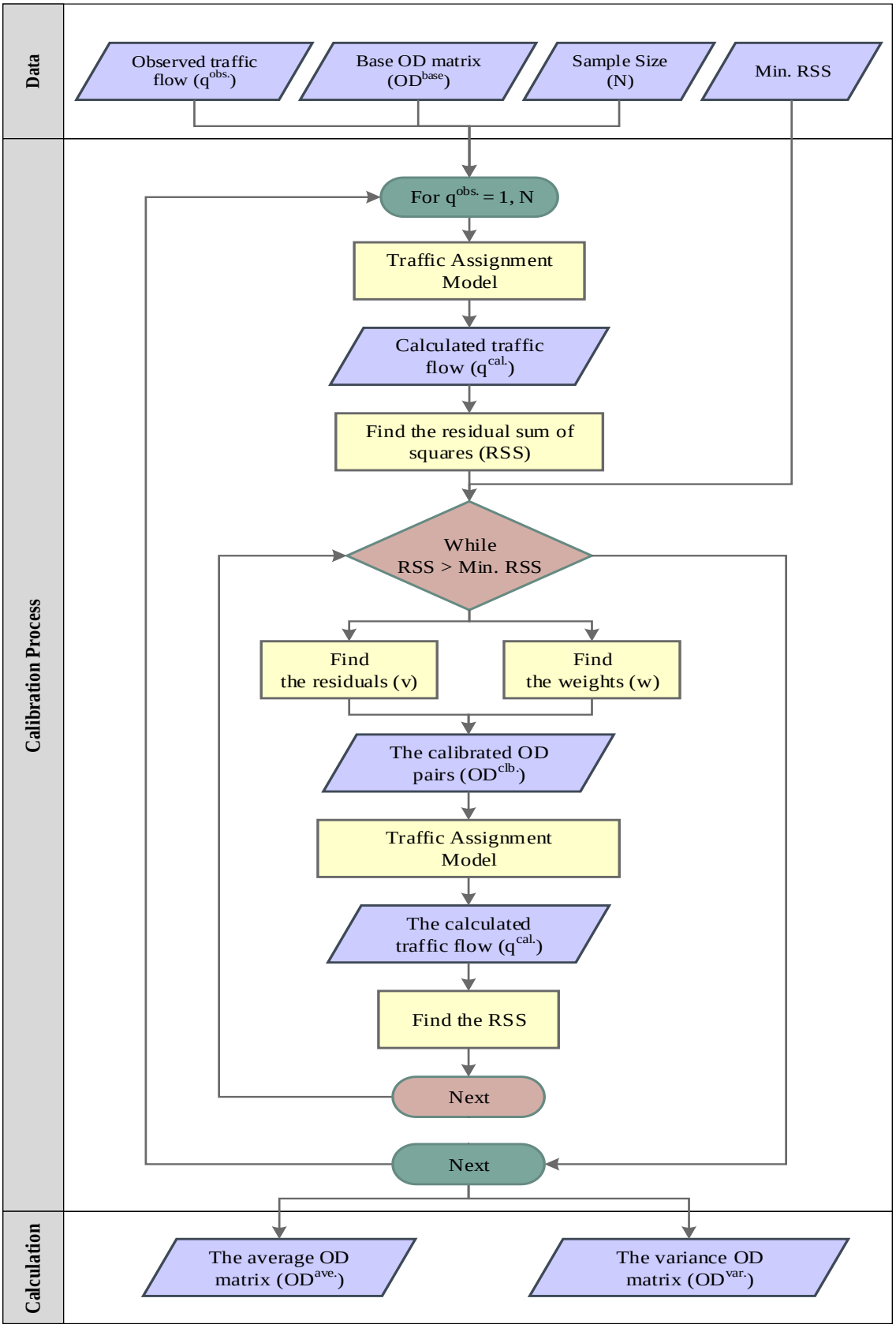

Figure 7: A flowchart illustrates the methodology stages and the relationship between the components of the stage.

Thesis V:

I have developed a new method capable of quantifying traffic flow uncertainty in the traffic assignment model by adding the variance parameters to the OD matrix.

This methodology was built using Forward Uncertainty Quantification (Forward-UQ) technique by using both the average-OD matrix and variance-OD matrix as input data in the traffic assignment model. Figure 8 shows the flowchart of the Forward-UQ method. The result of this methodology enables us to quantify the uncertainty of predictive traffic flow under different confidence intervals, such as 68%, 95%, or 99%.

Figure 9 shows the difference between the traditional deterministic approach and the uncertainty quantification approach. The deterministic approach where each input variable has a singular value only, and the output of the traffic assignment model will be a single value. While the uncertainty quantification approach takes the distributions of each input variable into account, and the output of the traffic assignment model becomes a range of possible values.

For example, the max. probable traffic flow value of at $CI = 95\%$:

$$q_l^{max.prob.@CI=95\%} = f_{T_{O\to D}}\left(OD^{max.prob.@CI=95\%}\right) \quad (10)$$

And, the min. probable traffic flow value of at $CI = 95\%$:

$$q_l^{min.prob.@CI=95\%} = f_{T_{O\to D}}\left(OD^{min.prob.@CI=95\%}\right) \quad (11)$$

Where: $f_{T_{O\to D}}$ is a traffic assignment model.

$$OD^{max.prob.@CI=95\%} = OD^{ave.} + 1.96 \times \sqrt{OD^{var.}}$$

$$OD^{min.prob.@CI=95\%} = OD^{ave.} - 1.96 \times \sqrt{OD^{var.}}$$

Related publication: [J-3]

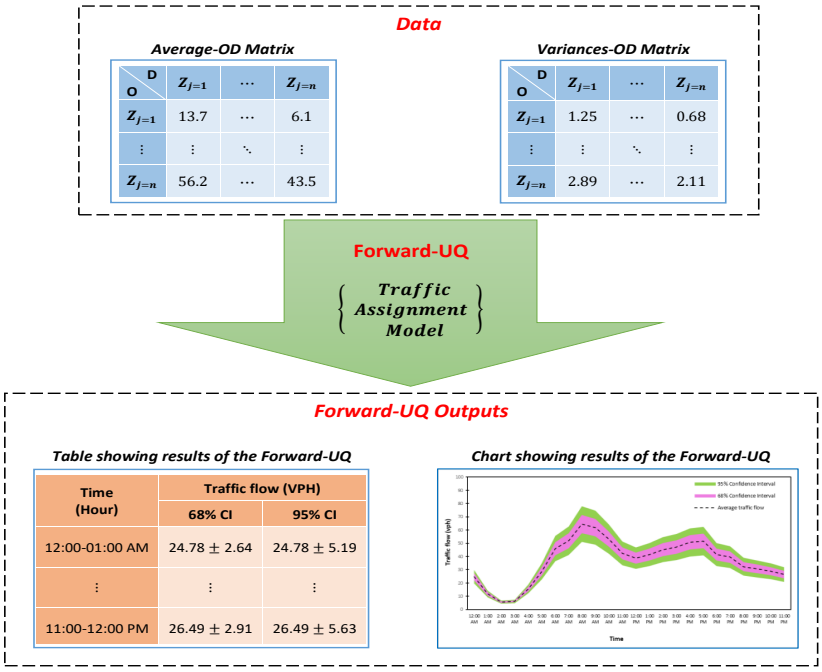

Figure 8: Flowchart of Forward-UQ method.

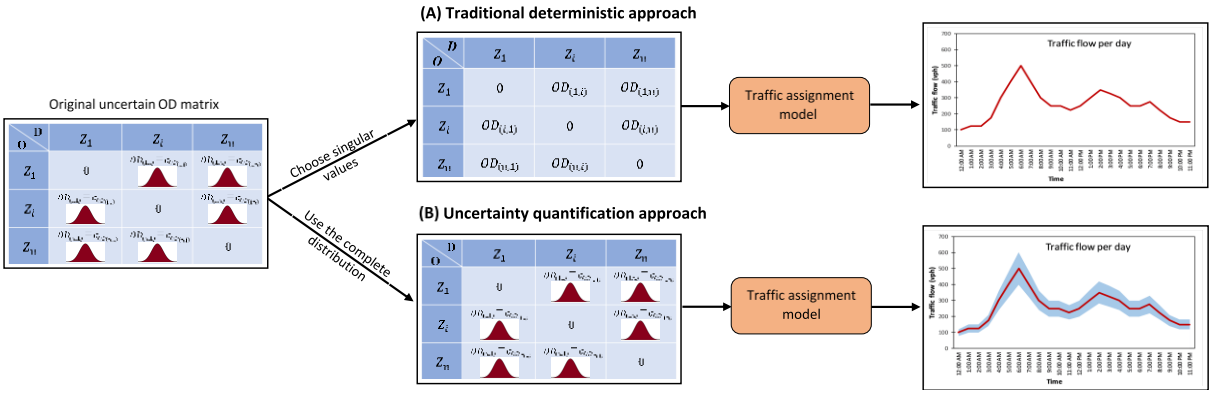

Figure 9: Shows the difference between the traditional deterministic approach vs the uncertainty quantification approach.

4. Outlook and Future Research

In the present dissertation, which took an in-depth search of the uncertainty analysis of the traffic assignment model caused by the data variation of the OD matrix, and proposed a novel methodology for quantifying the uncertainty of the traffic flow on the traffic assignment model. The developed methodology of uncertainty quantification on the traffic assignment model provides a practical, structured procedure for transportation administration organisations to calibrate the average-OD matrix and estimate the variance-OD matrix to predict the uncertainty of the traffic flow in the future.

However, as with any other research, recommendations, and suggestions for further research, improvement and refinement of the proposed methodology will improve its implementation and application in the transportation administration organisations.

Although the research's scientific results are important but more important, these results provide a basis for future research to extend the understanding of the subject. Thus, the following are some primary directions in which future work needs to be considered based on the presented study:

- In the direction of uncertainty analysis, the methodology of predicting traffic flow should consider applying different probability distributions varying according to travel purposes and land use characteristics.
- A continuous process of identifying and processing the QMC simulation sample size of the algorithm (GSA-TAM) needs to be carried out for every new case study separately.
- The methodology of link choice proportions can be used to build a new route choice model, besides comparing the new model results with other route choice models.
- The estimated proportions of the link choice methodology need to more research in aspects of detecting the probability distribution and statistical parameters in each link choice set.
- In the direction of uncertainty analysis, the proposed methodology handled mean and variance parameters to characterise the OD pairs' distribution; this is justified

for normal distribution only. Therefore, to enhance the quality of uncertainty quantification, other distributions should be tested through gaining higher-order statistics parameters.

5. Research Applications

The proposed framework provides a practical, structured procedure for analysing and quantifying the traffic flow uncertainty on the traffic assignment model. The framework will help transportation modellers and decision-makers have a fresh look at the uncertainty of the traffic assignments model's outputs. Figure 10 shows the uncertainty analysis and quantification framework.
The framework could be used for the following purposes:

- The major challenge of the OD matrix estimation problem is that it is severely under-determined. The proposed Uncertainty quantification approach can be used to estimate a more precise OD matrix.
- Enhance the traffic flows' prediction accuracy, which is particularly helpful to the Transport Management Systems (TMS).
- Monitor and identify which of the links that suffer from bias and unexpected change in traffic volumes in the event of a change in inputs OD matrix conditions.
- Risk analysis and management for the transport projects, especially the PPP (Public-Private Partnership) projects.

List of publications related to the thesis

[J-1] Seger, M., Kisgyörgy, L. "Predicting and Visualising the Uncertainty Propagations in Traffic Assignments Model Using Monte Carlo Simulation Method". Journal of Advanced Transportation, 2018. 2018. DOI:10.1155/2018/9825327

[J-2] Seger, M. A., Kisgyörgy, L. "Estimation of link choice probabilities using monte Carlo simulation and maximum likelihood estimation method." Periodica Polytechnica Civil Engineering, 64(1), pp. 20–32. 2020. DOI:10.3311/PPci.14366

[J-3] Seger, M., Kisgyörgy, L. "Uncertainty quantification of the traffic assignment model". Periodica Polytechnica Civil Engineering, 64(4), pp. 1181–1201. 2020. DOI:10.3311/PPci.16369

[J-4] Seger, M., Kisgyörgy, L. "Sensitivity analysis of input data of traffic assignment model". [under review]

[C-1] Seger, M., Kisgyörgy, L. "Visualise the Effect of Input Variability on Model Output in Traffic Assignment". In S. Lakušić (Ed.), Road and Rail Infrastructure V, Proceedings of the Conference CETRA (pp. 1233–1240). Zadar, Croatia: Department of Transportation Faculty of Civil Engineering University of Zagreb. 2018. DOI:10.5592/CO/CETRA.2018.683

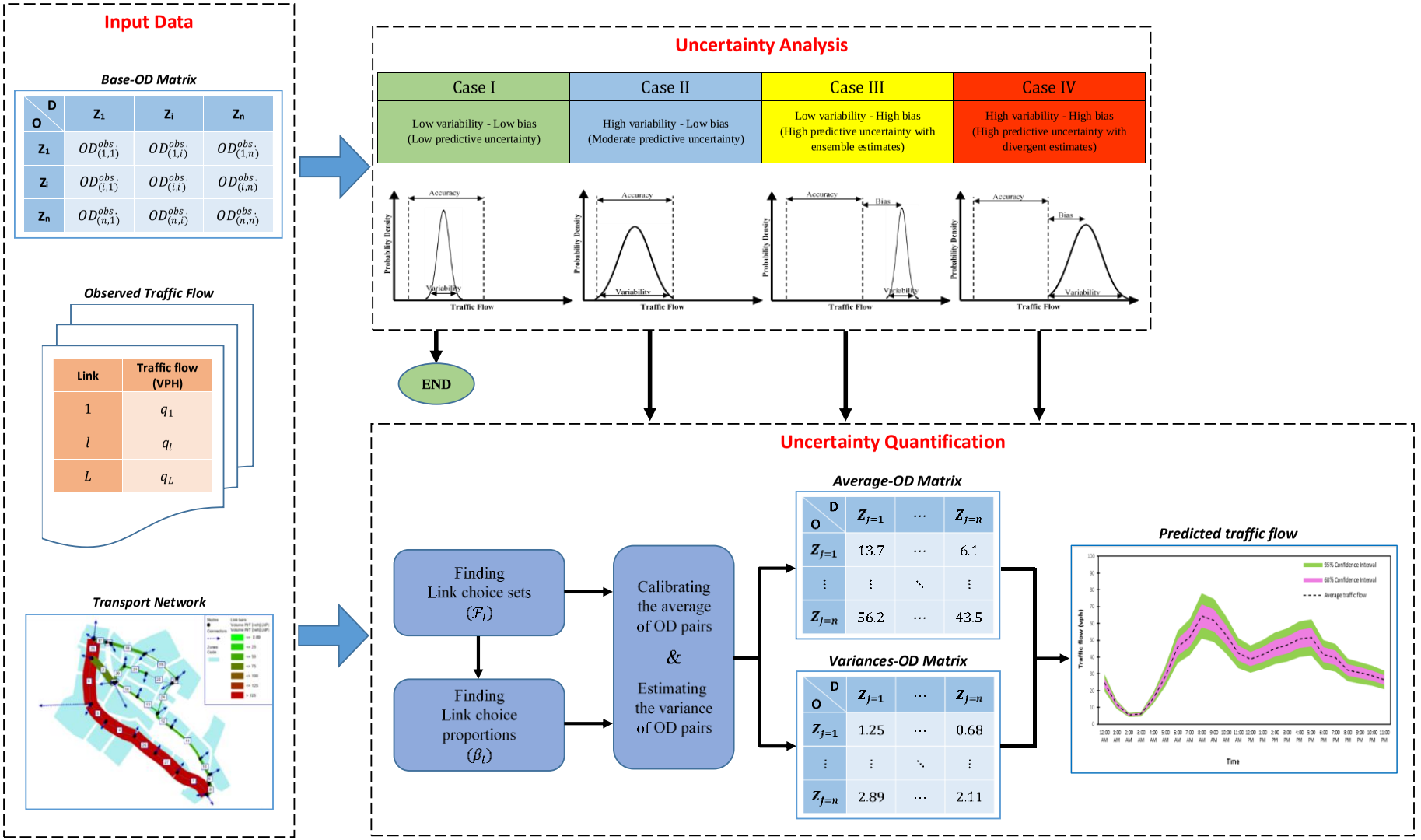

Figure 10: The uncertainty analysis and quantification framework.