MŰEGYETEM 1782

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

ANALYSIS AND QUANTIFICATION OF UNCERTAINTY IN THE TRAFFIC ASSIGNMENT MODEL

PhD dissertation

MUNDHER ALI SEGER

Supervisor:

LAJOS KISGYÖRGY, PhD

Associate Professor

Budapest, 2020

Declaration of Authenticity

I declare that all materials presented in this dissertation are either my own work or fully and specifically acknowledged wherever adapted from other sources.

Budapest, October 20th, 2020

Mundher Ali Seger

Acknowledgements

I would like to express my special gratitude to my supervisor Dr. Lajos Kisgyörgy, for all of his help, continuous support and advice throughout my research.

I would equally like to thank all the patience and support of my family especially my wife Esraa and my kids Zahraa, Mohammed and Mawadda, who provided me with a stable environment to make focusing on the work possible, also sacrificed a large amount of time I had to spend on this research.

Finally, I would also like to thank all of the colleagues at the Department of Highway and Railway Engineering who were always there to help when needed.

Table of Contents

List of Abbreviations and Symbols

Abbreviations

CI	Confidence intervals
COM	Component Object Model
FAST	Fourier Amplitude Sensitivity Test
Foreword-UQ	Foreword Uncertainty Quantification
GEH	Geoffrey E. Havers statistical test
GP	Gaussian Process
GSA	Global Sensitivity Analysis
GSA-TAM	Global Sensitivity Analysis-Traffic Assignment Model
GOF	Goodness-Of-Fit
Inverse-UQ	Inverse Uncertainty Quantification
KS test	Kolmogorov–Smirnov test
LS	Least Squares
LSA	Least Squares Adjustment
MC	Monte Carlo
MLE	Maximum Likelihood Estimation
OD	Origin Destination
PD	Probability Distribution
PRCC	Partial Rank Correlation Coefficient
QMC	Quasi-Monte Carlo
RSD	Relative Standard Deviations
RSS	Residual Sum of Squares
SA	Sensitivity Analysis
SI	Sensitivity Indices
TAZ	Traffic Analysis Zones
TDM	Travel Demand Models
TN	Transport Network Traffic
TSI	Total sensitivity indices matrix,
UQ	Uncertainty quantification

Greek Symbols

β_l	Link choice probabilities set for the link (l)
$\widehat{\beta_l}$	Estimated link choice probabilities set for the link (l)
η_l	Bias in traffic volume for the link (l)
θ	Parameter
θ_0	Unknown parameter
$\hat{\theta}_N$	Maximum likelihood estimator
Θ	Parameter space
λ^l	Correction factor of each OD pair in a link (l)
ε_l	Error in (q_l)
$\widehat{\varepsilon_l}$	Estimated error in (q_l)
μ	Mean parameter of PD
σ_l	Standard deviation of predicted traffic flow in a link (l)
${\sigma^2}_l$	Errors' variance of predicted traffic flow in a link (l)
$\widehat{{\sigma^2}_l}$	Estimated errors' variance of predicted traffic flow in a link (l)
$\{\varphi\}$	Sample value using the Monte Carlo procedure

Υ	Vector for the model residuals
v_l	Residual of the traffic flow in a link (l)
Ω^p	Representing the p-dimensional unit hyperspace

Roman Symbols

C	Choice set
D-	Measures the largest leak in observation compared with theoretical
D+	Measures the largest excess in observation compared with theoretical
D	Measures the largest difference in between
$f_{T_{O \to D}}$	Traffic assignment model
f_0	Constant in the variance-based sensitivity analysis method
G	All the non-empty subsets of the master set (M)
k	Iteration sample
KS	Critical value obtained from a table of the KS test
l	Link
L	Total number of links
M	Master set of known routes
M_1, M_2, M_3	Input matrices generated using QMC and OD matrix
N	Number of samples/iterations/observations
OD^{base}	Base-OD matrix
$OD^{ave.}$	Average-OD matrix
$OD^{var.}$	Variance-OD matrix
$OD^{clb.}$	Calibrated OD matrix
$OD^{max.pob.}$	Maximum probable-OD matrix
$OD^{min.pob.}$	Minimum probable-OD matrix
p	Number of parameters/variables in a model
$P(r)$	Probability of a driver choosing route
$P(r\|C)$	Probability of a driver is choosing route (r) from a given choice set (C)
$P(C)$	Probability of a driver's choice set being (C)
q_l	Traffic flow in a link (l)
$q_{l=1,L}$	predicted traffic flow attribute,
$\widehat{q}_l$	Estimated traffic flow in a link (l)
$\bar{q}_l$	Average predicted traffic flow in a link (l)
$q_l^{obs.}$	Observed traffic flow in a link (l)
q_l^{low}	Lower accurate limit in a link (l)
q_l^{up}	Upper accurate limit in a link (l)
$q_l^{cal.}$	Calculated traffic flow in a link (l)
$q_l^{est.}$	Estimated traffic flow in a link (l)
$q_l^{obs.}$	Observed traffic flow in a link (l)
Q_1, Q_2, Q_3	Traffic flow attributes corresponding to input matrices M_1, M_2, M_3
r	Choosing route
S_i	First-order sensitivity index
S_{ij}	Second-order index
ST_i	Total sensitivity index
U	Universal set that includes all possible routes for an OD pair
V	Variance
W	Set of the weighing for the parameters/variables
w^l	Weight of each OD pair in every link

X	Set of input parameters/variables
X^c	Set of the correction for parameters/variables
X°	Set of approximate values of the unknown parameters/variables
Y	Set of model outputs
$Y^{mod.}$	Vector for the model outputs
$Y^{obs.}$	Vector for the observations

Abstract

Traffic assignment models play a significant role in the transportation planning process. However, uncertainty in traffic flow is the most critical matters facing the traffic assignment models. Uncertainty is a well-acknowledged phenomenon that occupies interest the transport planners and decision-makers alike increasingly in recent years. Uncertainty in traffic assignment models mainly arises from many input variables that essentially come from the Origin-Destination (OD) matrix. Therefore, to increase the traffic assignment model's reliability, it is necessary to identify, calibrate, and estimate the statistical parameters that describe the OD pairs' variances, which helps analyze and quantify the traffic assignment model's uncertainty.

Consequently, the dissertation introduces a newly developed practical framework consisting of two approaches; the first approach is the uncertainty analysis, and the second approach is the uncertainty quantification. The developed framework provides a practical, structured procedure for transportation organizations to enhance traffic assignment models' prediction accuracy.

In the uncertainty analysis approach: a new methodology has been developed to calculate the observed traffic flows' uncertainty. This methodology is based on the Monte Carlo (MC) simulation method associated with the probability distributions of the OD matrix elements. The traffic assignment model's uncertainty was displayed in multiple prediction scenarios and quantified in terms of bias and variance. The importance of this methodology is permitting transport planners and decision-makers to monitor and identify which of the links that suffer from bias and unexpected change in traffic volumes in the event of a change in the conditions of inputs OD matrix.

For the uncertainty quantification approach, a novel methodology has been innovated to quantify traffic flow uncertainty in the traffic assignment model. This methodology consists of four subsequent steps: In the first step, a new method was built to identify OD pairs' choice sets that create the traffic flow on all links in the transport network. This method depends on the global sensitivity analysis of the variance-based technique and quasi-Monte Carlo simulation. The second step aims to find the choice set proportions. Thus, a new method has been developed based on the Maximum Likelihood Estimation (MLE) and Gaussian Process (GP) to estimate the link choice proportion for each OD pair

in every link separately. In the third step, an innovative method was created to determine all OD pairs' distribution parameters in the OD matrix. This method is built depending on the Inverse Uncertainty Quantification (Inverse-UQ) technique through Least Square Adjustment (LSE) method. This method produces two matrices instead of the one input OD matrix: (1) The average-OD matrix, (2) The variance-OD matrix. The final step includes a new methodology using Forward Uncertainty Quantification (Forward-UQ) technique by inputting both the average-OD and variance-OD matrices in the traffic assignment model. This methodology's importance is permitting transport planners and decision-makers to expect the uncertainty of the predictive traffic flow in different levels of confidence.

1. Introduction

1.1. 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 categorized 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 [1–5]. With well-known previous studies in mind, which is show a lack of knowledge in analyzing 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 analyzing 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.

1.2. 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 analyzing and quantifying the traffic flows' uncertainty. The challenges here are: define the necessary information that useful in analyzing 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 analyzing and quantifying traffic flows uncertainty in traffic assignment model.

1.3. Data

The data used in this study was collected from a small city called Ajka, located in Hungary. Figure (1) shows the Transport network (TN) and traffic analysis zones (TAZ) of the study area. The TAZ consists of 25 zones (22 internal zones and 3 external zones), and 25 two-directions links (i.e., 50 links). This study's required data consists of the base OD matrix and traffic flow in all links for a duration of time; the more extended time duration, the results will be more accurate.

The details of the transport network can be found in Appendix (A).

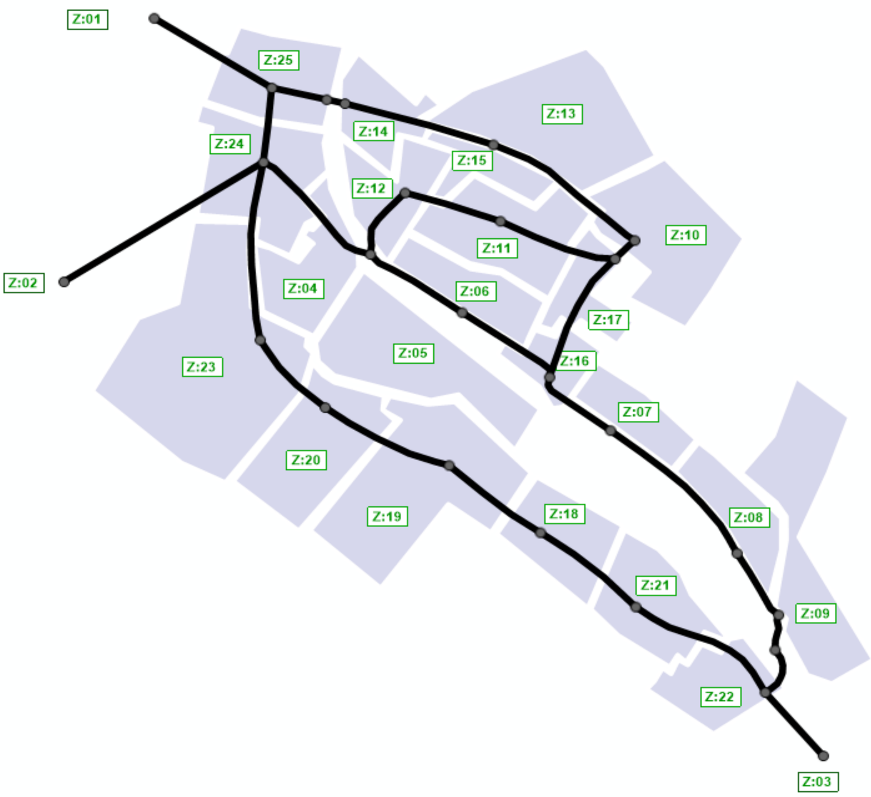


Figure 1: Study area (Ajka, Hungary), showing the Traffic Analysis Zones (TAZ) and transport network (TN).

1.4. Dissertation Layout

In order to meet the study's objectives, it was found necessary to carry the research in several stages. The stages are described in six chapters. The following paragraphs explain the core contents of each chapter in this dissertation.

Chapter 1: Provides the background to the issues in uncertainty in the traffic assignment model, the purpose and motivation behind this research, the key questions and the objectives of this research, and the data used for executing the proposed methodologies in this dissertation.

Chapter 2: Titled "Predicting, classifying and visualizing the uncertainty propagation in traffic assignments model using Monte Carlo simulation method," presents a methodology to predict the traffic flow uncertainty caused by the stochastic variation of the OD matrix.

Chapter 3: Titled " Sensitivity analysis of input data of traffic assignment model," presents a methodology 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.

Chapter 4: Titled "Estimation of link choice probabilities using maximum likelihood estimation method and Gaussian processes," presents a methodology to estimate choice proportions for each link on the transport network.

Chapter 5: Titled " Uncertainty quantification of the traffic assignment model." First presents a methodology to determine the distribution parameters for all pairs in the OD matrix from the traffic flow data using the inverse uncertainty quantification approach. Next, it presents a methodology to quantify traffic flows uncertainty in the traffic assignment model using the forward uncertainty quantification approach.

Chapter 6: Summarizes this dissertation's achievements and new scientific results, and presents some major directions for future research that enhance the traffic assignment model's uncertainty quantification.

2. Predicting, Classifying and Visualizing the Uncertainty Propagation in Traffic Assignments Model Using Monte Carlo Simulation Method

2.1. Introduction

Forecasting of travel demand represents the fundamental step of planning and management of transportation facilities [6]. These forecasts are subjected to various sources of errors: error in the measurement of input data, error in the estimated value of model parameters, error in the specification of the underlying models themselves; also, the model itself may be stochastic, and the scenarios adopted for model forecasting may not necessarily be compatible with the real evolution of the transport system [2, 3]. Studying the uncertainty of travel demand is not only aimed at stripping the values of traffic flow from errors but mainly aims to know the effect of these errors regarding the nature of the variability and bias on traffic flow and the likelihood of occurrences; this will be discussed in this study.

The main goal of the travel demand model is traffic forecasting in different stages; generation, distribution, and assignment are to determine future values of the model output variables that are associated with a specific combination of input variables[1, 4]. However, it is impossible to give an exact prediction; no model can be constructed to provide 100% accurate predictions of a system's future behaviour. A prediction should handle uncertainties by treating output variables stochastically. Without the additional information provided by probability analysis, there is no concrete evidence for comparing the predicted value to real value or another prediction [5, 6]. As a result, any method used for prediction should include assessing the uncertainty in the predicted values. Uncertainty in traffic forecasts is inherent significantly in travel demand models. Accordingly, a solid understanding of the degree of uncertainty helpful not only for the model designation process but also for the sampling process and the value of precise model results for policy scenario development and strategy evaluation [3, 7, 8].

Although many researchers have studied uncertainty in the travel demand models, only a few studies analyzed the error propagation impact in whole four-stages sequential transport model frameworks. Such as, Zhao and Kockelman [2] found that the uncertainty increases throughout the first three model steps, and declined in the traffic assignment step. They argued that this reduction might be due to the network congestion effects on the trip assignment equilibrium system, meaning that capacity limitations might reduce the variability of the link traffic flows' outcomes. However, they also looked out that reducing uncertainty in the traffic assignment might also result from the accumulation of the same links of independent trips related to various OD pairs. De Jong et al. [6, 10] had examined the Dutch national model's uncertainty using the standard deviations and correlations of 20-year of some input data; they found that congestion reduced the final model output uncertainty but only to a minor degree. Ziems et al. [15] analyzed the effects of randomness on different traffic characteristics; They observed that traffic flow variability is larger during the day's congested period and greater variability in the individual corridor rather than aggregated subareas. Manzo et al. [12, 13] stated that transport network congestion does not show a high uncertainty effect on the transport model's final output. Hence, the final uncertainty of traffic flow for links with a higher volume/capacity ratio showed a lower dispersion around the base uncertainty value. Rasouli and Timmermans [14,15] investigated the OD matrix's uncertainty using the Dutch national transport model. They found that higher levels of traffic flow result in lower levels of uncertainty for different model outputs. Thus, the researcher emphasized that the degree of uncertainty grows higher if the focus of attention shifts from aggregate system performance indicators through the OD matrix to disaggregated space-time sequences and performance indicators.

Future year travel demand forecasting is not an exact science, and there are complicated underlying mechanisms that inherently generate uncertainty in the forecasts. Modelling these complicated mechanisms requires numerous variables and behavioural components whose variability may be poorly determined or ignored. In this case, it is illogical to take a single view of the future without considering the uncertainty in travel demand modelling. Thus, to provide more efficient and reliable transport solutions for the future, transport analysts and planners must observe and predict uncertainty in transport systems [3, 8, 16].

Uncertainty becomes relevant in transportation modelling in the case of diverging views such as; if risks are very high if the policy is controversial and concerns about model limitations. However, in a certain view, several points estimate based on different scenarios are given to account for uncertainty [20]. Ideally, analysts would wish to understand the separate and collective impact of these errors on the uncertainty of model forecasts, to be able to attach credible confidence intervals to model forecasts and optimize the allocation of study resources [13]. However, in large model systems, the interaction between these error sources can be very complicated, analyzing uncertainty propagation extremely challenging through the modelling process [18, 19]. Nevertheless, the private sector's increased participation in the delivery of transport infrastructure projects has raised the requirement for accurate traffic demand forecasts and led to renewed interest in the analysis of model uncertainty [23].

The traffic assignment stage is the most crucial stage of travel demand modelling. The traffic assignment model considers the dependencies among origin-destination demand, link flows, and path costs. Additionally, simulate the interactions between transport system supply and travel demand to achieve an output that explains a transport system's general state and overall variability [24]. The traffic assignment model outcomes are generally essential to decisions related to infrastructure development and transport policy measures. Therefore, simple point estimates of the link flow, even if it refers to the most probable values, is not enough for an accurate and safe assessment of the risks associated with those decisions. Consequently, the need for quantifying more accurately the uncertainty in traffic assignment estimates is strongly present [25, 26].

This chapter presents a new methodology for exploring the nature of uncertainty propagation deriving from the input OD matrix in a traffic assignment model. This methodology consists of three parts; the first part predicts the traffic flows establishing from a specific probability distribution, standard deviation, and base OD matrix; after that, compares predicted values to real values to explore the traffic flow uncertainty. The second part classifies the traffic flows uncertainties into four cases according to variability and biases in traffic assignment results. Finally, is visualizing the uncertainty results in charts and figures. This methodology's product assists the transport-modellers and decision-makers to have a fresh look at the effect of the input data (i.e., OD matrix) on the traffic assignments models' uncertainty.

2.2. Methodology

A new methodology was developed for analyzing and characterizing the uncertainty of the observed traffic flows. This work's structure directly supports a visual segmentation of uncertainty for the transport network to present error and bias in traffic volumes calculated by traffic assignment models. This methodology consists of six stages:

(i) Input stage,

(ii) MC simulation process stage,

(iii) Analysis of predicted traffic flow stage,

(iv) Predictive uncertainty stage,

(v) Uncertainty classification stage,

(vi) Uncertainty visualization stage.

The mathematical and logical computations of this methodology were illustrated in Algorithm 1, and the relationships connecting the stages are presented in Figure 2.

Algorithm 1: Prediction of Uncertainty in Traffic Assignment using Monte Carlo Simulation

---

Require: $Z_{j=1,n}$; $\forall Z_j \in \mathbb{N}$ ⊳ Zone definition.
$O_{j=1,n}$; $O_j \subset Z$ ⊳ Origin definition.
$D_{j=1,n}$; $D_j \subset Z$ ⊳ Destination definition.
n; $\forall n \in \mathbb{N}$ ⊳ number of zones.
$OD_{(n,n)}$; $\forall OD_{(n,n)}$ ⊳ Matrix of Origin Destination (OD).
$OD^{base}_{(j,j)}$; $\forall OD^{base}_{(j,j)} \in \mathbb{R}$ ⊳ base value of OD pair.
$f_{T_{O \to D}}$ ⊳ Traffic Assignment function.
$q^{obs.}_{l=1,L}$; $\forall\, q^{obs.}_l \in \mathbb{R}$ ⊳ Observed values of traffic assignment function.
L; $\forall\, L \in \mathbb{N}$ ⊳ Number of Links of the transport network.
PD ⊳ Define the Probability Distribution type.
μ, RSD; $\forall\, \mu, \sigma \in \mathbb{R}$ ⊳ Parameters range for the probability distribution.
N ; $\forall N \in \mathbb{N}$ ⊳ Number of samples.

$\{\varphi_k\} \leftarrow MC\ [PD\ (\mu, RSD)]$ ⊳ Get samples by the defined probability distribution and parameters using Monte Carlo procedure.
For $k = 1\ to\ N$ do
  For $j = 1\ to\ n$ do
    For $j = 1\ to\ n$ do
      $OD^k_{(j,j)} \leftarrow OD^{base}_{(j,j)} * \{\varphi_k\}$ ⊳ Get the generated-OD matrices.
    end for
  end for
end for
For $k = 1\ to\ N$ do ⊳ Run Visum.
  $q^k_{l=1,L} \leftarrow f_{T_{O \to D}}(OD^k_{(j,j)})$; $\forall\, q^k_l \in \mathbb{R}$ ⊳ Get the predicted traffic flow attribute.
end for
For $l = 1\ to\ L$ do
  $\bar{q}_l \leftarrow \frac{\sum_{k=1}^{N} q^k_l}{N}$ ⊳ Get the average predicted traffic flow in each link.
  $\sigma_l \leftarrow \sqrt{\frac{1}{N}\sum_{k=1}^{N}(q^k_l - \bar{q}_l)^2}$; $\forall\, \sigma_l \in \mathbb{R}$ ⊳ Get the standard deviation in each link.
  $\eta_l \leftarrow (\bar{q}_l - q^{obs.}_l)$; $\forall \eta_l \in \mathbb{R}$ ⊳ Get the total bias for each link.
  $q^{low}_l \leftarrow \bar{q}_l + 6.25 + 1.25\sqrt{16\ \bar{q}_l + 25}$ ⊳ Get the lower accurate limit for each link.
  $q^{up}_l \leftarrow \bar{q}_l + 6.25 - 1.25\sqrt{16\ \bar{q}_l + 25}$ ⊳ Get the upper accurate limit for each link.
end for
For $l = 1\ to\ L$ do
  For $k = 1\ to\ N$ do ⊳ Check the variety and the bias for each sample.
    If $\bar{q}_l - \sigma_l \le q^k_l \le \bar{q}_l + \sigma_l$ Then ⊳ Low-Variety
    Else ⊳ High-Variety
    If $q^{low}_l \le q^{pre.}_l \le\ q^{up}_l$ Then ⊳ Low-Bias
    Else ⊳ High-Bias
  end for
end for
End

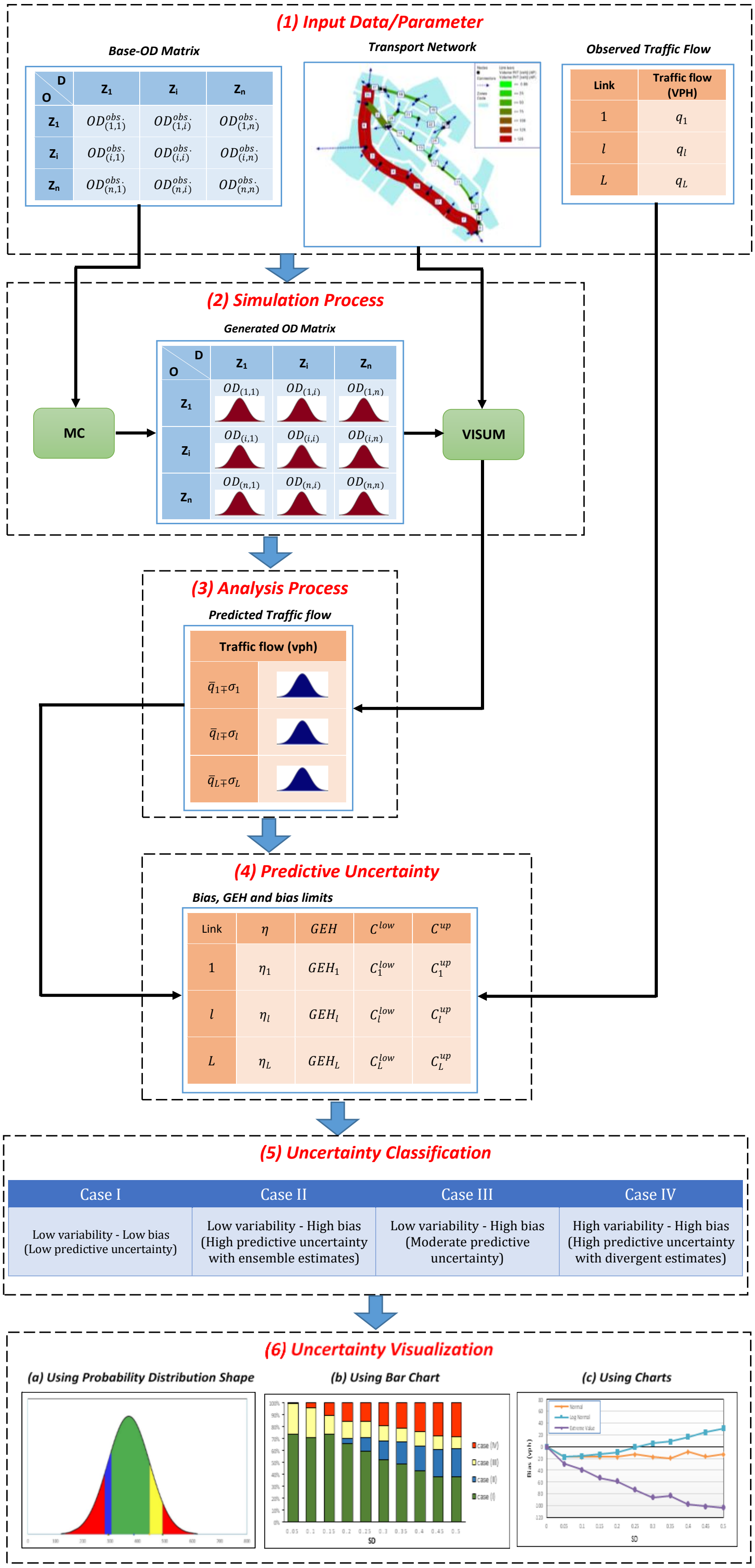


Figure 2: Flowchart of the methodology of uncertainty prediction, characterization, and visualization.

2.2.1. Input Stage

The principal task in predictive modelling is to estimate the behaviour of a modelling function, in this case, the traffic assignment function $f_{T_{O \to D}}$defined to calculate the traffic flow between Origin zone (O) to Destination zone (D). This work addresses the case where $f_{T_{O \to D}}$ can be calculated at a finite set of iterations (N). Monte Carlo simulation method was used to generate the input data with the parameters; relative standard deviation (RSD) and mean (μ) to produce finite OD matrices. The required data consist of three parts: the first part is setting physical features of the Transport Network (TN) and Traffic Analysis Zones (TAZ) by VISUM. The second part is defining the observed OD matrix. Finally, the third part finds the observed traffic flow (q_l^{obs}) by counting the real value of traffic flows. Monte Carlo simulation method is used to generate the OD matrices. Therefore, the required parameters for this process are; relative standard deviation (RSD), mean (μ), and the number of iterations (N). Moreover, the type of probability distribution (PD) must be chosen at the beginning. The required input data and parameters are explained in Algorithm 1.

2.2.2. Monte Carlo Simulation Stage

Monte Carlo (MC) method or random sampling method is a division of computational mathematics. It created from the mathematics concepts for (the frequency approximates the probability). When the solution for a problem is the occurrence probability of a certain event or an expected value of any variants, they use as a testing method to obtain the occurrence frequency of an event or the average value for these variants. MC method is based on the probability model and according to the described process by this model. The results of the simulation test are approximate solutions [27]. MC method plays a fundamental role in the characterization and quantification of uncertainty. When the accurate calculation of output uncertainties needed, then Monte Carlo based analysis is a reliable technique, and it is widely applicable. As a result, its application can be found in virtually all engineering fields. Monte Carlo simulation was usually utilized to observe how a system's errors or variability can propagate to the final result [28].

In this stage, the simulation code has been written using both MATLAB and Component Object Model (COM) of VISUM software to predict traffic flows on the transport network. The simulation code involves two processes: the first is the generation of the OD matrix

(OD^k) by Monte Carlo method for a given type of probability distribution via base-OD matrix (OD^{base}).

$$\{\varphi_k\}_{k=1}^{N} = MC\ [PD\ (\mu, RSD)] \tag{1}$$

$$OD^k = OD^{base} * \{\varphi_k\} \tag{2}$$

Where:

$\{\varphi_k\}$	sample value using Monte Carlo procedure,
MC	Monte Carlo procedure,
k	iteration number, (N) is number of iterations,
PD	probability distribution,
μ	mean parameter of PD,
RSD	relative standard deviation parameter of PD,
OD^k	generated OD matrix,
OD^{base}	base-OD matrix.

While the second process is the process of entering the generated OD matrix; and the Transport Network (TN) in the VISUM to produce the traffic flow attribute for the transport links from the traffic assignment function.

$$q_{l=1,L}^{k} = f_{T_{O\to D}}\ (OD^k) \tag{3}$$

Where:

$f_{T_{O\to D}}$	traffic assignment model,
$q_{l=1,L}^{k}$	predicted traffic flow attribute,
l	link number,
L	number of links.

This simulation is repeated for a finite number of iterations (N) specified by the researcher. With a note, an increasing number of iterations leads to improve the accuracy of results but extending simulation time.

The details of Visum software via COM Interface Programming can be found in Appendix (B).

2.2.3. Analysis of Predicted Traffic Flow Stage

One of the most important aspects of a simulation study is an analysis of simulation experiments. In this stage, the simulation process results are analyzed for statistical characterization. The outcomes of the simulation process are traffic flow attributes for all links on the transport network. The number of obtained attributes is equal to the number of iterations (N) used in the simulation.

The statistical characteristics of predictive traffic flow are defined by two parameters; the average value results ($\bar{q}_l$), and the variability (σ_l). In this case, the average traffic flow for all links in the transport network as Equation 4. Finally, the standard deviation represents the variability of results as Equation 5.

$$\bar{q}_l = \sum_{k=1}^{\mathrm{N}} q_l^k / N \tag{4}$$

$$\sigma_l = \sqrt{\frac{1}{N} \sum (q_l^k - \bar{q}_l)^2} \tag{5}$$

Where:

q_l^k predicted traffic flow in a link (l),

$\bar{q}_l$ average predicted traffic flow in a link (l),

σ_l the standard deviation of predicted traffic flow in a link (l).

2.2.4. Predictive Uncertainty Stage

Modelling transport systems commonly subject to uncertainty comes mainly from the input data used to build the OD matrix. This data is characterized as heterogeneous, insufficient, unstable, and unsteady because of the difficulty measurements. The nature of individual behaviour leads to variate the elements of the OD matrix. Besides limitations and lack of information in mathematical concepts and the structural in four-stages sequential transport model frameworks leads to a variate of traffic flow in transport networks.

The predictive uncertainty is defined by joint consideration of the mean predictive error (i.e., statistical bias) and the predictive variability (i.e., statistical standard deviation). In this methodology, the uncertainty has been calculated by joining the observed traffic flow attribute from the input stage (first stage), and the traffic flow attributes obtained by simulation processes (fourth stage).

The mathematical operations in this stage include calculating bias of the observed traffic flows from the average predictive traffic flow ($\bar{q}_l$); as Equation 6 and determining the limitations of allowed biases in the transport network.

$$\eta_l = (\bar{q}_l - q_l^{obs.}) \tag{6}$$

Where:

η_l bias in traffic volume for the link (l),

$q_l^{obs.}$ The observed traffic flow for the link (l).

In any uncertainty quantification process, setting limits for the predictive uncertainty is necessary to increase the researchers' understanding of the model's behaviour in both bias and predictive 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 7 [6]. It is helpful for these comparisons because it is sophisticated of relative and absolute errors. Moreover, as the standard error estimation, the standard deviation statistic was adopted as a limitation for traffic volumes variability.

$$GEH = \sqrt{\frac{2\,(q_l - \bar{q}_l)^2}{(q_l + \bar{q}_l)}} \tag{7}$$

GEH statistic band less than 5 [6] is used to explain the bias limit for each link.

By solving Equation 7, the lower and the upper bias limit in predicted traffic flow as follows:

$$q_l^{low} = \bar{q}_l + 6.25 - 1.25 \times \sqrt{16\,\bar{q}_l + 25} \tag{8}$$

$$q_l^{up} = \bar{q}_l + 6.25 + 1.25 \times \sqrt{16\,\bar{q}_l + 25} \tag{9}$$

Where:

q_l^{low} the lower accuracy limit for the link (l),

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

The allowed accuracy and variability are presenting in Equation 10 and Equation 11, respectively.

$$Accuracy\ allowable\ for\ the\ link\ (l) = q_l^{up} - q_l^{low} \tag{10}$$

$$Variblity\ allowable\ for\ the\ link\ (l) = \bar{q}_l \mp \sigma_l \tag{11}$$

If the observed traffic flow values lying within the standard deviation, then these are precise. Otherwise, these predicted values are imprecise (Equation 12 and Equation 13).

$$\text{If } \bar{q}_l - \sigma_l \le q_l^{obs} \le \bar{q}_l + \sigma_l, \text{ then low variety (precise)} \quad (12)$$

$$\text{If } \bar{q}_l - \sigma_l > q_l^{obs} < \bar{q}_l + \sigma_l, \text{ then high variety (imprecise)} \quad (13)$$

In the same way, if the observed traffic flow values lying within upper and lower bias limits, then these are accurate. Otherwise, these predicted values are biased (Equation 14 and Equation 15).

$$\text{If } q_l^{low} \le q_l^{obs} \le q_l^{up}, \text{ then low bias (accurate)} \quad (14)$$

$$\text{If } q_l^{low} > q_l^{obs} < q_l^{up}, \text{ then high bias (inaccurate)} \quad (15)$$

To interpret the concept of precision (i.e., the variability of values) and the accuracy (i.e., the bias of values) (see Figure 3).

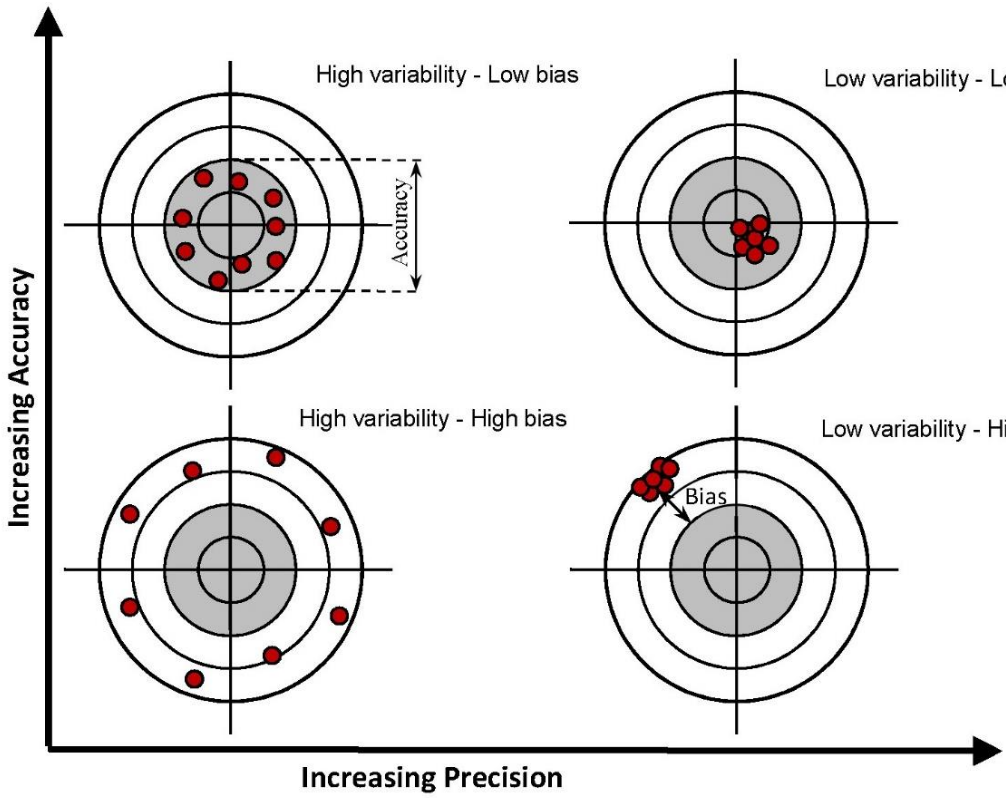


Figure 3: The relationship between precision and accuracy.

2.2.5. Classifying Predictive Uncertainty Stage

In this stage, a systematic classification is proposed for the predictive uncertainty of the observed traffic flows. The proposed classification has subdivided into four cases according to bias and variability:

Case I: Accurate and precise (Low predictive uncertainty). This case occurred when the observed traffic flows (q_l^{obs}) are close to the average predictive traffic flow ($\bar{q}_l$), i.e., within the standard deviation ($\pm\sigma_l$). And the observed traffic flows are within the accuracy limits (q_l^{low} and q_l^{up}). That is means the results are low variability and low bias.

Case II: Accurate and imprecise (Moderate uncertainty). This case occurred when the observed traffic flows (q_l^{obs}) are close to the average predictive traffic flow ($\bar{q}_l$), i.e., within the standard deviation ($\pm\sigma_l$). And the observed traffic flows are out of the accuracy limits (q_l^{low} and q_l^{up}). That is means the results are low variability and high bias.

Case III: Inaccurate and precise (High predictive uncertainty with ensemble agreement). This case occurred when the observed traffic flows (q_l^{obs}) are far from the average predictive traffic flow ($\bar{q}_l$), i.e., out the standard deviation ($\pm\sigma_l$). And the observed traffic flows are within the accuracy limits (q_l^{low} and q_l^{up}). That is means the results are high variability and low bias.

Case IV: Inaccurate and imprecise (High predictive uncertainty with divergent estimates). This case occurred when the observed traffic flows (q_l^{obs}) are far from the average predictive traffic flow ($\bar{q}_l$), i.e., out the standard deviation ($\pm\sigma_l$). And the observed traffic flows (q_l^{obs}) are out of the accuracy limits (q_l^{low} and q_l^{up}). That is means the results are high variability and high bias.

Figure 4 demonstrates the probability density curves of predictive uncertainty cases.

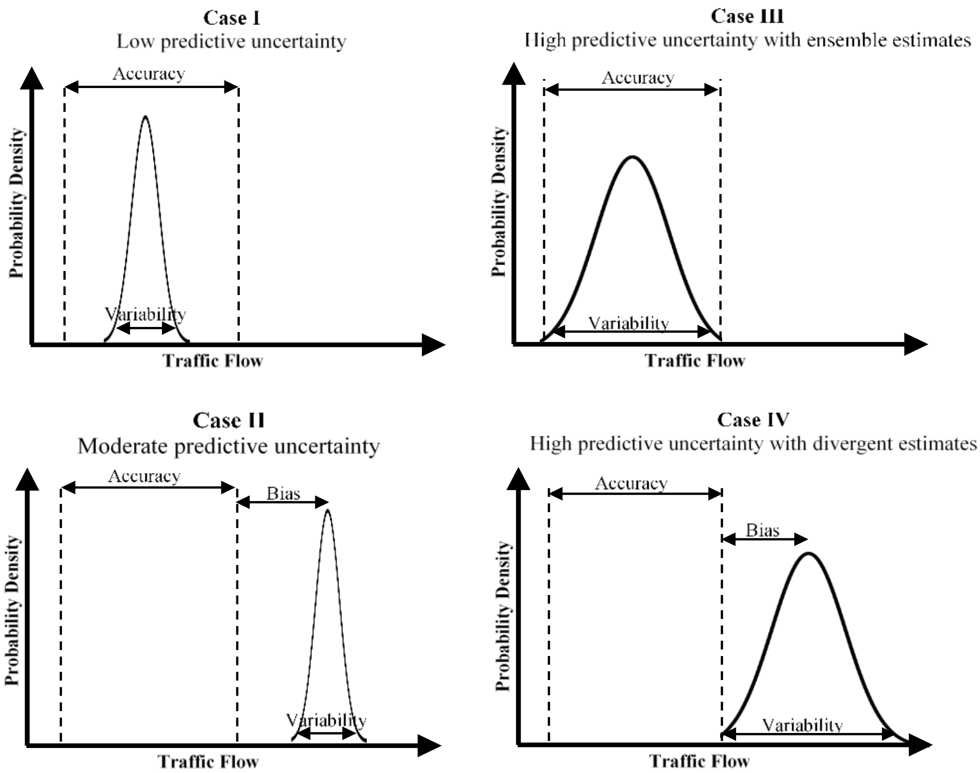


Figure 4: Presents the probability density curves of the predicted traffic flow for the four cases.

2.2.6. Visualizing Predictive Uncertainty Stage

The last stage of this methodology is uncertainty visualization. Uncertainty visualization is endeavouring to display data together with additional uncertainty information. These visualizations present a more complete and accurate interpretation of data for researchers to analyze [29]. Thus, visualization is a useful method for addressing many forms of information uncertainty, and it is a helpful approach to the investigation and communicating of large data sets [30].

Applications that use visual graphs and comparative figures to indicate information variability or draw confidence levels in data values help analysts better understand and cope with uncertain information better than using digital tables and metadata [31]. Consequently, visualizing uncertainty is essential for risk analysis and decision-making tasks. However, it is still a challenge because describing the uncertainty is a complex concept with many interactions, definitions, and interpretations in transportation models. Uncertainty can be introduced into information visualizations as the data is collected, transformed, and integrated into information [25, 26]. In the absence of the combined presentation of data and its associated uncertainty, the information visualization analysis is incomplete at best, and it may lead to inaccurate or incorrect conclusions. Therefore, there is a need to display information and their uncertainty for accurate interpretation and precise decision making [26, 28, 29].

Different methods are used to visualize uncertainty; statistical and probability-based visualization, point and global visualization used colours, financial visualization, icons, ontology, and lexicon [23]. In this methodology, the statistical and probability-based visualization method is used to visualize uncertainty. This method is one of the most powerful methods to address conceptual model uncertainty is with a bar chart, probability distributions, and traditional charts represented by random variables. This method demonstrates the central tendency, dispersion, skewness, and modal characteristics of a random variable. In this methodology, two steps are using to visualize the uncertainty of the predicted traffic flow.

The primary role of uncertainty visualization is giving information about the level of uncertainty to the decision-makers. Based on these cases, they can see how reliable the model's predictions are and if they can accept the risks relating to the given predictive uncertainty. If predictions that are more precise needed, sensitivity analysis helps identify the input data most dominantly influencing the predictive uncertainty. Then these input data should be measured more precisely. Therefore, a specific colour has been given for each case of uncertainty. Table 1 illustrates the colours, characterizations, and limitations of the four uncertainty cases. In practice, all these four cases can appear with different likelihood proportions; thus, Figure 5 shows an example of a probability density curve for observed traffic flow for a link (l). The coloured areas under the curve represent the predictive uncertainty cases.

Table 1: This table shows the classification of the four cases of uncertainties

		Variability limitations (Precision)	
		Low	High
Bias limitations (Accuracy)	Low	**Case I** Low variability - Low bias (Low predictive uncertainty)	**Case II** High variability - Low bias (Moderate predictive uncertainty)
	High	**Case III** Low variability - High bias (High predictive uncertainty with ensemble estimates)	**Case IV** High variability - High bias (High predictive uncertainty with divergent estimates)

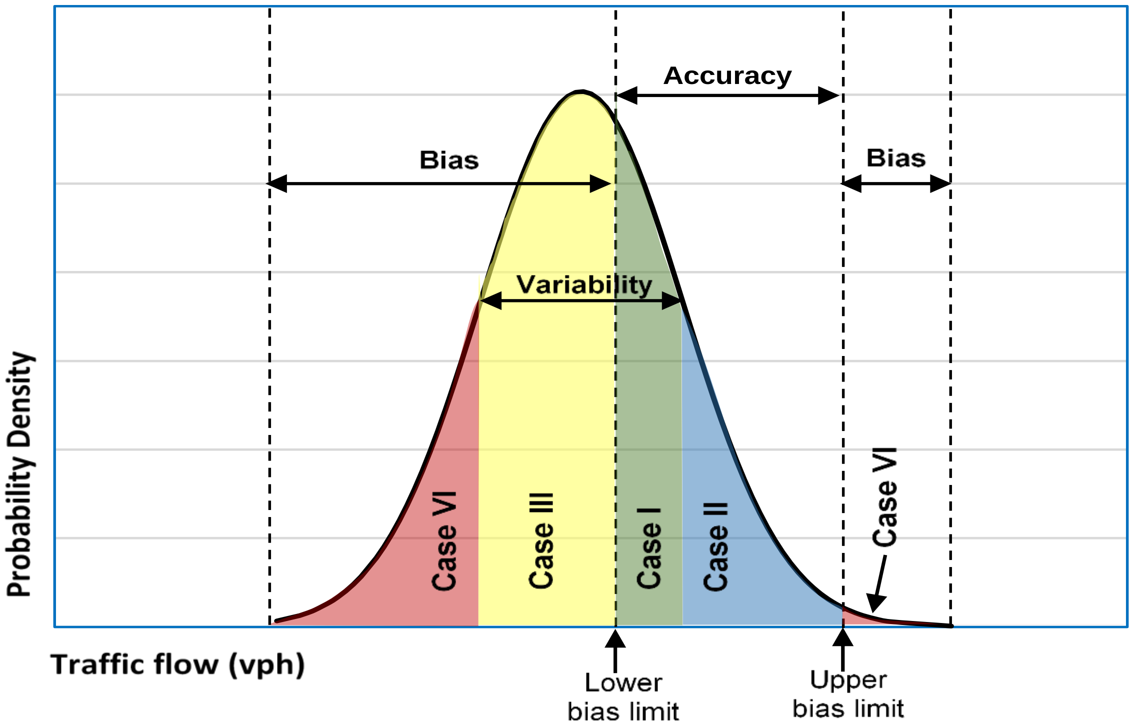


Figure 5: Shows the probability density of the observed traffic flows addressed with the uncertainty cases.

2.3. Implementation

The proposed methodology has been applied to the case study (Ajka, Hungary). The following subsections represent the implementation stages:

2.3.1 Input Data and Parameters

The required data consist of the base-OD matrix (OD^{base}), observed traffic flow ($q_{l=1,L}^{obs.}$) and type of probability distributions (PD) with the relative standard deviation value (RSD).

In this study, Normal probability distribution with (RSD = 0.2) and base OD matrix (OD^{base}) have been used to generate predictive traffic flow (q_l). And because of the fewness of the observed traffic flows data, three types of probability distributions (PD): (Normal Distribution, Lognormal Distribution, Extreme value distribution) with ten values of relative standard deviations (RSD) ranged from 0.05 to 0.50 have been used to generate observed traffic flows (q_l^{obs}).

2.3.2 Monte Carlo Simulation

In this simulation, 30 scenarios have experimented. These scenarios have been grouped into three groups according to the probability distribution (PD) type. Furthermore, for each group, the parameter of relative standard deviations (RSD) has been changed for each scenario. Besides, 1000 samples have been used per each scenario. Thus, the total number of samples used in the whole simulation process: (30 scenarios) x (1000 samples for each scenario) equal to (30000 samples). Therefore, this simulation is expensive; the time spent to complete this simulation is around (300 hours).

The simulation time of this methodology depends on; the number of links on the (TN), the number of (TAZ), and the number of samples (N).

2.3.3. Analysis of Predicted Traffic Flow

The simulation process results are represented by one attribute of average predictive traffic flow (q_l). And, thirty attributes of observed traffic flow (q_l^{obs}). Consequently, the outcomes for each link have been studied separately by finding the statistical parameters for predictive traffic flows; both of the average value ($\bar{q}_l$), and the variability (σ_l). Because there are 50 links on the transport network in this study case, only link #6 (l_{06}) is presented, demonstrating the methodology.

2.3.4. Predictive Uncertainty

The obtained data from the previous stage has been processed by Excel-sheet to find the uncertainty in traffic flows. The mathematical calculations determine the observed traffic flows' bias and variability and estimate accurate limits in each link separately. For (l_{06}): the lower accurate limit (q_{06}^{low}) = 295 vph, and the upper accurate limit (q_{06}^{up}) = 492 vph.

Figure 6 shows the percentage of bias in the observed traffic flows associated with input relative standard deviations (RSD). It can be observed that by increasing the relative standard deviation of inputs leads to increase bias for the observed traffic flows, for normal distribution scenarios and lognormal distribution scenarios; there are no or few biases for $RSD \leq 0.10$, and the biases increased to reach about 42% in $RSD = 0.50$. While for extreme value distribution scenarios, the biases increasing in all standard deviations to reach about 54% in $RSD = 0.50$.

Figure 7 shows the percentage of converging degree for observed traffic flows joined with input RSD It can be noted that the variability the of observed traffic flows for normal distribution scenarios ranged between 67.1% to 70.3% for $RSD \leq 0.50$ and decreasing after that to reach to 57.6% in $RSD = 0.50$. Moreover, the variability of observed traffic flows for lognormal distribution scenarios ranged between 67.9% to 70.3% for $RSD = 0.40$ and increasing after that to reach to 76.4% in the $RSD = 0.50$. While the variability of observed traffic flows for extreme value distribution scenarios ranged between 67.2% to 72.0% for $RSD \leq 0.35$ and decreasing after that to reach 61.4% in the $RSD = 0.50$.

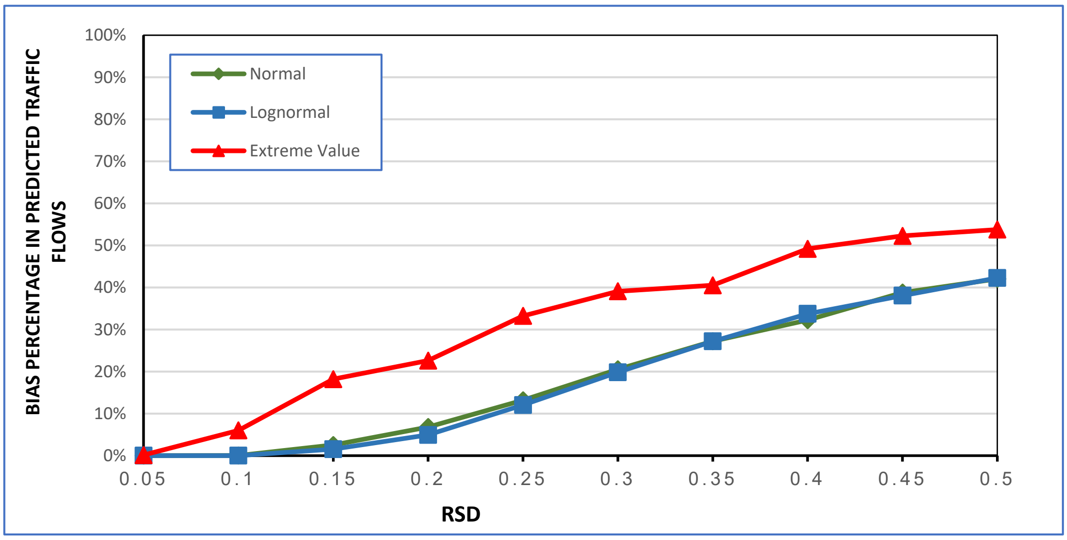


Figure 6: Shows the relationship between the relative standard deviation (RSD) for input OD matrices and the bias percentages of the observed traffic flows in (l_{06}).

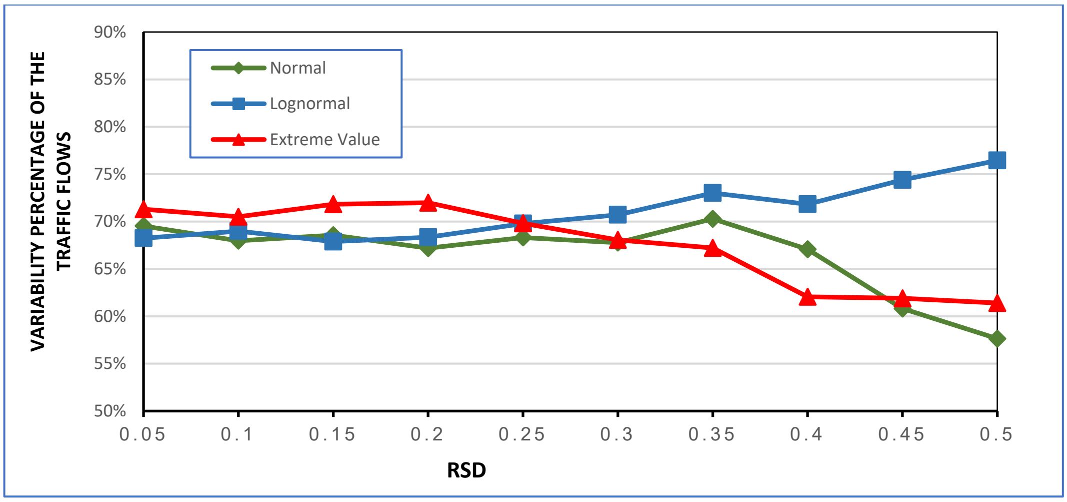


Figure 7: Shows the relationship between the relative standard deviation (RSD) for input OD matrices and the variability of the observed traffic flows in (l_{06}).

2.3.5. Classifying Predictive Uncertainty

The analysis stage's statistical results have been categorized and classified according to bias limits and variability into four cases. Hence, the classification is varied in each link according to the adopted scenario.

For example, in the scenario:

- $PD = Normal, RSD = 0.20$,
- The average observed traffic flow $\bar{q}_{06} = 391\ vph$,
- The lower accuracy limit $\left(q_{06}^{up} = 492\ vph\right)$ Equation (8),
- The upper accuracy limit $\left(q_{06}^{low} = 295\ vph\right)$ Equation (9),
- The standard deviation $(\sigma_{06} = \pm 50.6\ vph)$,
- The variability limit $(391 \pm 50.6 = 340.4\ vph, 441.6\ vph)$ Equation (11),

The classification result for this scenario is presented in Table 2.

Table 2. This table shows the classification of (l_{06}) in scenario: PD= Normal and RSD=0.20.

	$295 \leq q_{06} \leq 492$	$295 > q_{06} < 492$
$340.4 \leq q_{06} \leq 441.6$	**Case I** Low variability - Low bias	**Case II** Low variability - High bias
$340.4 > q_{06} < 441.6$	**Case III** High variability - Low bias	**Case IV** High variability - High bias

2.3.6. Visualizing Predictive Uncertainty

Finally, the probability of the uncertainty cases for the observed traffic flows in the transport network has been visualized into bar-charts by merging Figure 8, Figure 9, and Figure 10. This visualization provides transport planners and engineers with the possibility to monitor and identify which of the links that suffer from bias and unexpected changes in traffic volumes in the event of a change in traffic parameters and experiment with different scenarios.

The coloured bar-charts give a whole idea about the probability ratios of the uncertainty cases. In general, the percentage of green colour (Case I) means the probability of observed traffic flow will lay within the allowed limits of accuracy and variability. Moreover, the percentage of yellow colour (Case III) means the probability of observed traffic flow will lay still within the allowed accuracy but outer than the allowed variability. Both percentages of green colour (Case I) and yellow colour (Case III) located within the accuracy range. While the percentage of blue colour (Case II) means the probability of observed traffic flow will lay outer than the allowed accuracy but still within the allowed variability. The percentage of red colour (Case IV) means the probability of observed traffic flow will lay outside than the allowed limits of accuracy and variability. Both percentages of blue colour (Case II) and red colour (Case IV) are located outside the accuracy range.

Although the purpose of the research is visualization the uncertainty, it is good to give more details by text. In the same example, for (l_{06}): concerning normal distribution scenarios, as shown in Figure 8, which shows: (Case I) is decreased from 69.5% in RSD=0.05 to 54.7% in RSD=0.50. (Case II) start appearing from RSD=0.30 and increased to 5.3% in RSD=0.35 and decreased after reaching 2.9% in RSD=0.50. (Case III) is decreased from 32 % in RSD=0.05 to 3.2% in RSD=0.50. (Case IV) started appearing from RSD=0.15 and increased to reach 39.1% in RSD=0.50.

Similarly, Figure 9 can be interpreting for lognormal scenarios, which it shows: (Case I) is decreased from 68.2% in RSD=0.05 to 57.7% in RSD=0.50. (Case II) start appearing from RSD=0.35 and increased to 18.7% in RSD=0.50. (Case III) is decreased from 31.8% in RSD=0.05 to 0.6% in RSD=0.35 and disappear after that. (Case IV) started appearing from RSD=0.15 and increased to reach 23.6% in RSD=0.50.

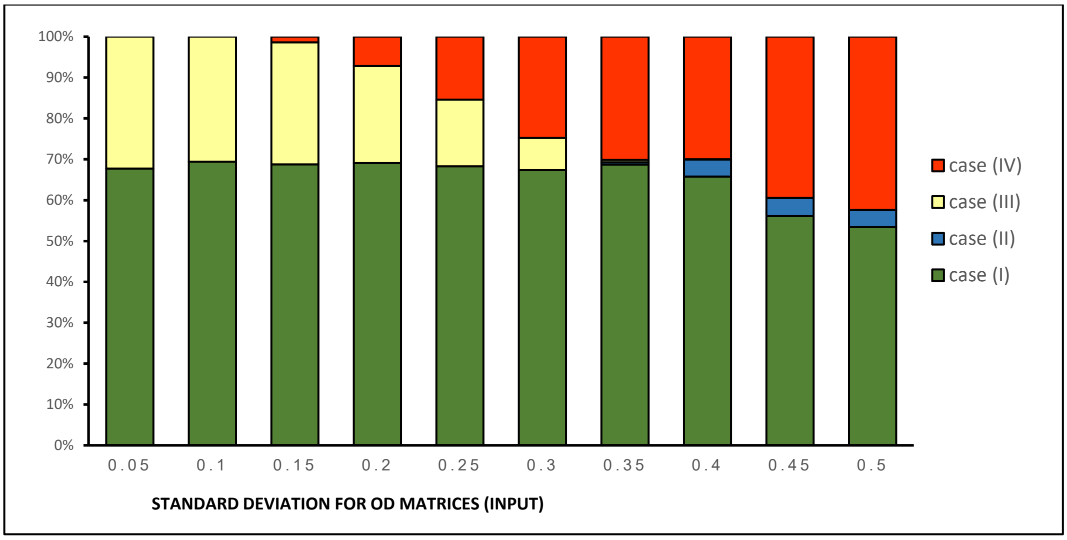


Figure 8: Represent the observed uncertainty for link #6 (l_{06}), applying normal distribution scenarios.

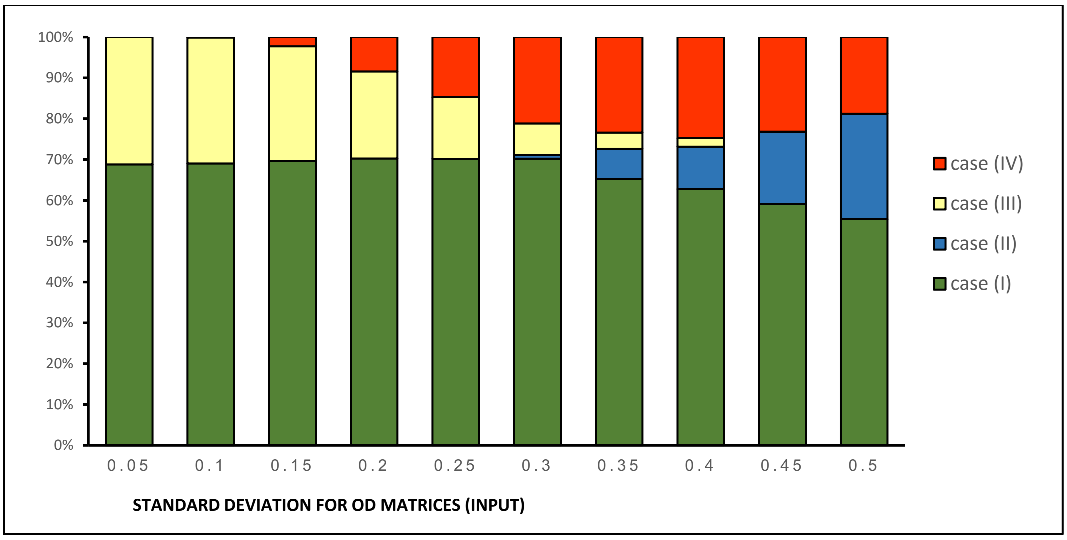


Figure 9: Represent the observed uncertainty for (l_{06}), applying lognormal distribution scenarios.

Likewise, Figure 10 can be interpreting for extreme value scenarios; which it shows: (Case I) is decreased from 71.3% in RSD=0.05 to 33.3% in RSD=0.50. (Case II) start appearing from RSD=0.15 and increased to 28.1% in RSD=0.50. (Case III) is decreased from 28.6% in RSD=0.05 to 12.6% in RSD=0.50. (Case IV) started increasing from 0.6% in RSD=0.05 to 25.7% in RSD=0.50.

Figure 8, Figure 9, and Figure 10 show that the uncertainty analysis results strongly depend on the uncertainties of the input data. That is proving that the application of stochastic approaches in the practice of transport modelling highly advisable. Thus, uncertainty analysis of the observed traffic flows indicates data acceptability and helps to identify the relating risks.

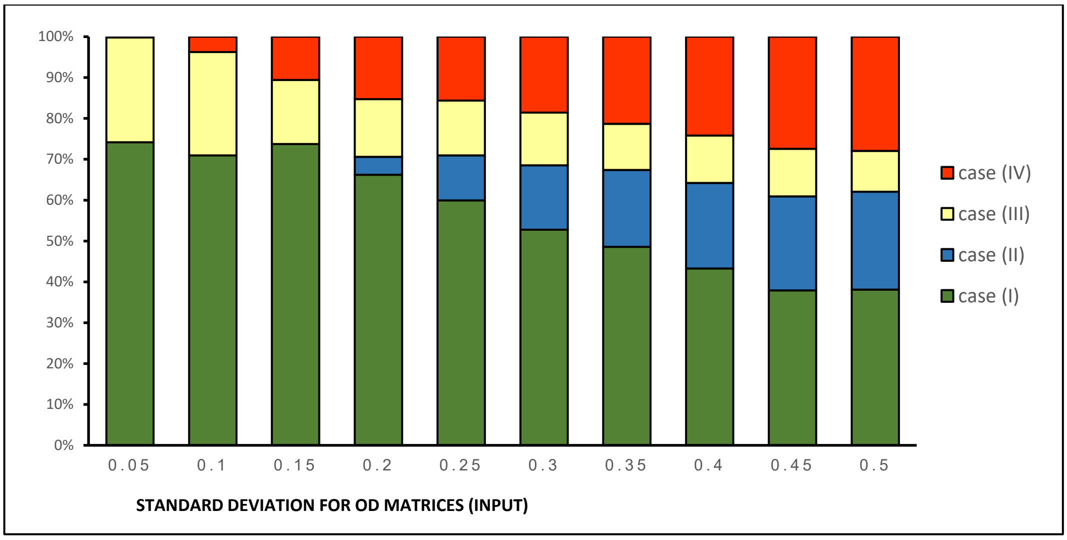


Figure 10: Represent the observed uncertainty for (l_{06}), applying extreme value distribution scenarios.

2.4. Conclusions

This chapter presents a new methodology to predict traffic flow and visualize the uncertainty in those observed values. This methodology enables transport planners to apply various scenarios to monitor the traffic flow variation on the transport network by supposing that the OD matrix's input values vary according to a specific probability distribution. The importance of this methodology is permitting transport planners and decision-makers to monitor and identify which of the links that suffer from bias and unexpected change in traffic flows in the event of a change in the conditions of inputs OD matrix.

This methodology's algorithm consists of two parts; the first part has been built on the MC simulation method to generate numerous OD matrices and VISUM software for getting the traffic assignment on a transport network. The results of this part represent predicted traffic flows on each link of the transport network. These predicted traffic flows are suffering from uncertainty in both a bias from the observed value and variability from the average predicted value. Furthermore, the second part of the algorithm was designed to categorize the uncertainty of the observed traffic flows into four cases according to variability and bias: Case I (low variability - low bias), Case II (low variability - high bias), Case III (high variability - low bias), and Case IV (high variability - high bias). Finally, the percentages of these cases have been visualized in coloured bar-charts. The percentage of each case represents the likelihood of occurring (i.e., the likelihood of the observed traffic flow to biasing or variating depends on these cases).

3. Sensitivity Analysis of Input Data of Traffic Assignment Model

3.1. Introduction

Sensitivity analysis (SA) permits identifying the parameter or set of parameters that significantly influence the model output. It gives useful insight into which of input variables contribute most to the variability of the model output [34, 35], consequently to obtain the most crucial input factors and the contribution of the variability input interactions. SA also can identify variables where the reduction in uncertainty will have a maximum contribution to increase model precision [36]. SA is used in various scientific applications; it can help understand the input-output relationship and identify the significant and influential parameters that raise model outputs and magnitudes. Moreover, the SA used to determine the uncertainty in setting the model structural parameters that contribute to the model output's overall variability, leading future experimental designs, and risk analysis [34, 37, 38].

In general, models are sensitive to input parameters in two distinct ways. First, the variability or uncertainty associated with a sensitive input parameter is propagated through the model resulting in a significant contribution to the overall output variability. Second, the model output can be positively correlated with input parameters so that small changes in the input values lead to significant changes in the output [39, 40]. Specification of the most influential factors on model outputs using SA techniques achieves valuable information for improving the model and identifying parameters that must be studied more carefully [41]. It will lead to an increase in the reliability and robustness of predictions obtained by large-scale models. SA is also useful for all processes where it is essential to know which input factors contribute the most to output variability. Sensitivity analysis is an effective way to quantitatively analyze output patterns' behaviour concerning perturbations of model inputs and parameters [12, 36].

Moreover, SA plays a significant role in engineering decisions under uncertainty by quantifying uncertainties on the output results of interest. For example, SA gives a ranking of the most influential uncertain input parameters. It provides insight into how reducing uncertainty in the inputs might correspondingly reduce uncertainty in the output results [23, 42, 43].

SA is usually classified into two categories: Local Sensitivity Analysis and Global Sensitivity Analysis. Local Sensitivity Analysis evaluates changes in the model outputs concerning variations in a single parameter input. In a Global Sensitivity Analysis (GSA), all parameters are changed simultaneously over each parameter space, which allows evaluating the relative contributions of each parameter and the interactions between parameters to the model output variance [38, 44, 45].

The most common GSA methods used are regression-based methods with rank transformation such as the Partial Rank Correlation Coefficient (PRCC) and variance-based methods such as the Sobol' method and Fourier Amplitude Sensitivity Test (FAST). These methods established methods are widely applied in a variety of engineering applications [42, 46].

Typically, SA is used to determine how "sensitive" a model changes in the value of the model's inputs/parameters and changes in the model's outputs. The distinction between the effect of the parameters and inputs is necessary to deep realizing and more understanding about model results' behaviour. However, various transportation researchers use SA in traffic assignment models such as Leurent [47], Waller et al. [25], Ceylan and Bell [48], Patriksson [49], Shao et al. [50], Kattan and Abdulhai [51], Sun et al. [52], Djukic [53], and Ma and Tang [54]. However, all these studies dealt with the parameter's sensitivity of model structure only. Therefore, there is a gap in knowledge about using the SA method to discover the influential and non-influential OD pairs in the traffic assignment model, examining the sensitivity of each link in the transport network to the OD's variability values matrix.

3.2. Global Sensitivity Analysis

Global sensitivity analysis (GSA) permits identifying the parameter or set of parameters that significantly influence the model's output variance [42]. GSA's objectives represent model verification, model simplifying, factors prioritization, system design safety justification, and the GSA to validate computer code and software [34]. GSA has been widely used in engineering design to get more information about complex model behaviour and support designers to make educated decisions about developing the engineering models [55]. One of the most widely GSA methods used in the engineering and applied sciences is variance-based sensitivity analysis, also known as analysis of variance (Sobol' method), based on the functional model decomposition introduced by the mathematician Sobol [56].

3.2.1. Variance-Based Sensitivity Analysis (Sobol method)

In GSA, variance-based sensitivity analysis or Sobol method refers to ways of quantifying the contribution of each input parameter to the total variance of the output [35, 56, 57]. The Sobol sensitivity indices are obtained from the analysis of variances ANOVA decomposition [56]. Sobol Method as a GSA tool [58, 57] can be summarized as a following:

Let model outputs (Y) is a function of parameters/variables $(X_1, X_2, X_3, \dots, X_p)$, and thus, this relation can be written as follows:

$$Y = f\left(X_1, X_2, X_3, \dots, X_p\right) \tag{16}$$

Where:

p the number of parameters/variables in the model.

In a sensitivity analysis, the researchers wish to know the amount of variance that would disappear if one or more input parameters were known and fixed [59].

Sobol' suggested to decompose the function f into summands of increasing dimensionality:

$$f\left(X_1, \dots, X_p\right) = f_0 + \sum_{i=1}^{p} f_i(X_i) + \sum_{i=1}^{p} \sum_{j=i+1}^{p} f_{i,j}\left(X_i, X_j\right) + \dots + f_{1,\dots,p}\left(X_1, \dots, X_p\right) \tag{17}$$

If the input factors are independent and each term in this Equation is chosen with zero average and is square-integrable, then f_0 is a constant: equal to the output's expectation value, and the summands are mutually orthogonal. Additionally, this decomposition is unique.

The total unconditional variance can be defined as:

$$V(Y) = \int_{\Omega^p} f^2(\boldsymbol{X})\, d\boldsymbol{X} - f_0 \tag{18}$$

with Ω^P representing the p-dimensional unit hyperspace (i.e., parameter ranges are scaled between 0 and 1). The partial variances, which are the total variance decomposition components, are computed from each of the terms in Equation (19).

$$V_{i_1 \dots i_s} = \int_0^1 \int_0^1 f_{i_1 \dots i_s}^2 \left(X_{i_1}, \dots, X_{i_s}\right) dX_{i_1} \dots dX_{i_s} \tag{19}$$

where $1 \le i_1 \le \dots \le i_s \le p$ and $s = 1, \dots, p$. With the assumption that the parameters are mutually orthogonal, this results in Equation (20) for the variance decomposition.

$$V(Y) = \sum_{i=1}^{p} V_i + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} V_{ij} + \dots + V_{1,\dots,p} \tag{20}$$

Where $\sum_i^p V_i$ is the sum of partial variances that involve the main effects of each input parameter, $\sum_i^p \sum_{j>i}^p V_{ij}$ Include each partial variances of the interaction of two input parameters, and so on. In this method, the variance contributions to the total output variance of individual parameters and parameter interactions can be determined.

The proportion of the partial variance identifies these contributions to the total variance. The Sobol sensitivity indices can be founded by dividing Equation 20 by total unconditional variance V(Y); it becomes:

$$1 = \sum_i^p S_i + \sum_i^p \sum_{j>i}^p S_{ij} + \cdots + S_{1,\ldots,p} \tag{21}$$

Where:

$\sum_i^p S_i$ the sum of the first-order indices,

$\sum_i^p \sum_{j>i}^p S_{ij}$ the sum of the second-order indices,

Consequently, the first order index (S_i) and the second order index (S_{ij}) for each input parameter are given as follow:

$$S_i = \frac{V_i}{V(Y)} \tag{22}$$

$$S_{ij} = \frac{V_{ij}}{V(Y)} \tag{23}$$

The first order index (S_i) is a measure for the variance contribution of the individual parameter (X_i) to the total model variance, and the impact on the model output variance of the interaction between parameters (X_i) and (X_j) is given by (S_{ij}). Finally, the total sensitivity index (ST_i) represent the result of the main effect of (X_i) and all its interactions with the other parameters (up to the p^{th} order). The calculation of (ST_i) can be calculated from the variance ($V_{\sim i}$) that results from the variation of all parameters, except (X_i).

$$ST_i = 1 - \frac{V_{\sim i}}{V(Y)} \tag{24}$$

By analyzing the difference between ST_i and S_i, one can determine the impact of the interactions between the input variable X_i and the other input variables. In general, both indices values provide adequate information to determine the analyzed model's sensitivity to individual input variables.

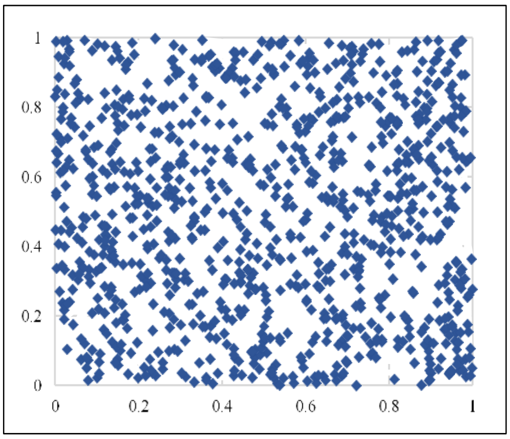
3.2.2. Quasi-Monte Carlo Algorithm Based on Sobol Sequences

3.2.2.1. Sobol sequences

Using the Quasi-Monte Carlo (QMC) simulation required a systematic sampling method to get deterministic inputs instead of random ones. In this chapter, the Sobol sequence has been used to generate a quasi-random of the number. In particular, it is common for QMC to produce much more accurate integration approximations answers than classical MC does [60]. Figure 11 illustrates the sampling difference between (a) classical MC and (b) QMC using Sobol' sequence.

In this study, the Sobol sequence has been used to produce random data for the simulation algorithm. This method generates a sequence of points in the unit hypercube $[0,1)^p$, where (p) is the dimension of the problem, i.e., the number of parameters that will be analyzed. In other words, each element inside the sequence is a p-dimensional vector whose components are fractions between 0 and 1.

(a) MC samples



(b) QMC samples using Sobol sequence

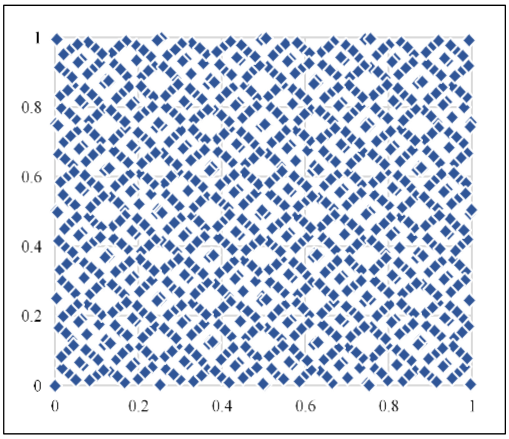


Figure 11: MC samples and QMC samples using the Sobol' sequence (N = 1000).

3.2.2.2. Quasi-Monte Carlo algorithm

The Sobol method is typically used to create two sets of input matrix $(N \times p)$, (the "sample" matrix M_1 and the "sample" matrix M_2).

Where:

N sample size,

p number of parameters/variables.

Every row in both matrices M_1 and M_2 represents possibly the parameter combination on the model [57, 61]. The first-order sensitivity index (S_i) and total sensitivity index (ST_i) were estimated by QMC algorithm:

$$S_i = \frac{\left[\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(y_k^{M_1}\right)^2 - f_0^2\right] - \left[\left(\frac{1}{2N}\right)\sum_{k=1}^{N}\left(y_k^{M_2} - y_k^{M_3}\right)^2\right]}{\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(y_k^{M_1}\right)^2 - f_0^2} \tag{25}$$

$$ST_i = \frac{\left(\frac{1}{2N}\right)\sum_{k=1}^{N}\left(y_k^{M_1} - y_k^{M_3 i}\right)^2}{\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(y_k^{M_1}\right)^2 - f_0^2} \tag{26}$$

Where $y_k^{M_1}$, $y_k^{M_2}$ and $y_k^{M_3}$ 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:

$$f_0^2 = \left(\frac{1}{N}\sum_{k=1}^{\mathrm{N}} y_k^{M_1}\right)^2 \tag{27}$$

3.3. Methodology

This chapter developed an algorithm of Global sensitivity analysis for the traffic assignment model (GSA-TAM) based on the QMC simulation technique. The GSA-TAM algorithm's primary purpose: First, identifying the set of OD pairs that creates traffic flow in each link of the transport network. Second, determine each OD pair's influence ratio 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 these stages are presented in Figure 12, and the mathematical and logical computations of this methodology were illustrated in Algorithm 2. The GSA-TAM algorithm stages are as follow:

1) **Definition stage**: in this stage, three things should be defined; (i) Travel demand model, the model might be a simple or complex according to the Number of TAZ, links, and nodes. (ii) Parameters (p), there are two types of parameters; the first one is the parameters associated with the model structure, while the second type is related to the input data. (iii) Sample size (N), the selection of sample size has a significant effect on the accuracy of sensitivity indices and time of the simulation

process. Increasing sample size leads to improved accuracy, but the simulation process will be expensive because of the time.

2) **Generation stage**: This stage includes building two sets of matrices using the QMC algorithm; in this study, the Sobol sequence was used. in both matrices (M_1) and (M_2), the number of columns represents the number of parameters, while the number of rows depicts the sample size.
3) **Running stage**: In this stage, the GSA-TAM algorithm has been coding by MATLAB software and linked with VISUM software via the Component Object Model (COM). The results of model running are two matrices of outputs (Q_1) and (Q_2), corresponding to the two sets of matrices (M_1) and (M_2) respectively.
4) **Calculation stage**: During this stage, both Equation 25 and Equation 26 are used to calculate the first-order sensitivity index and total sensitivity index, respectively.
5) **Analysis stage**: In the last stage, the set of OD pairs that affect each link will be identified. Moreover, determining the scope of effectiveness for each OD pair on the transport network.

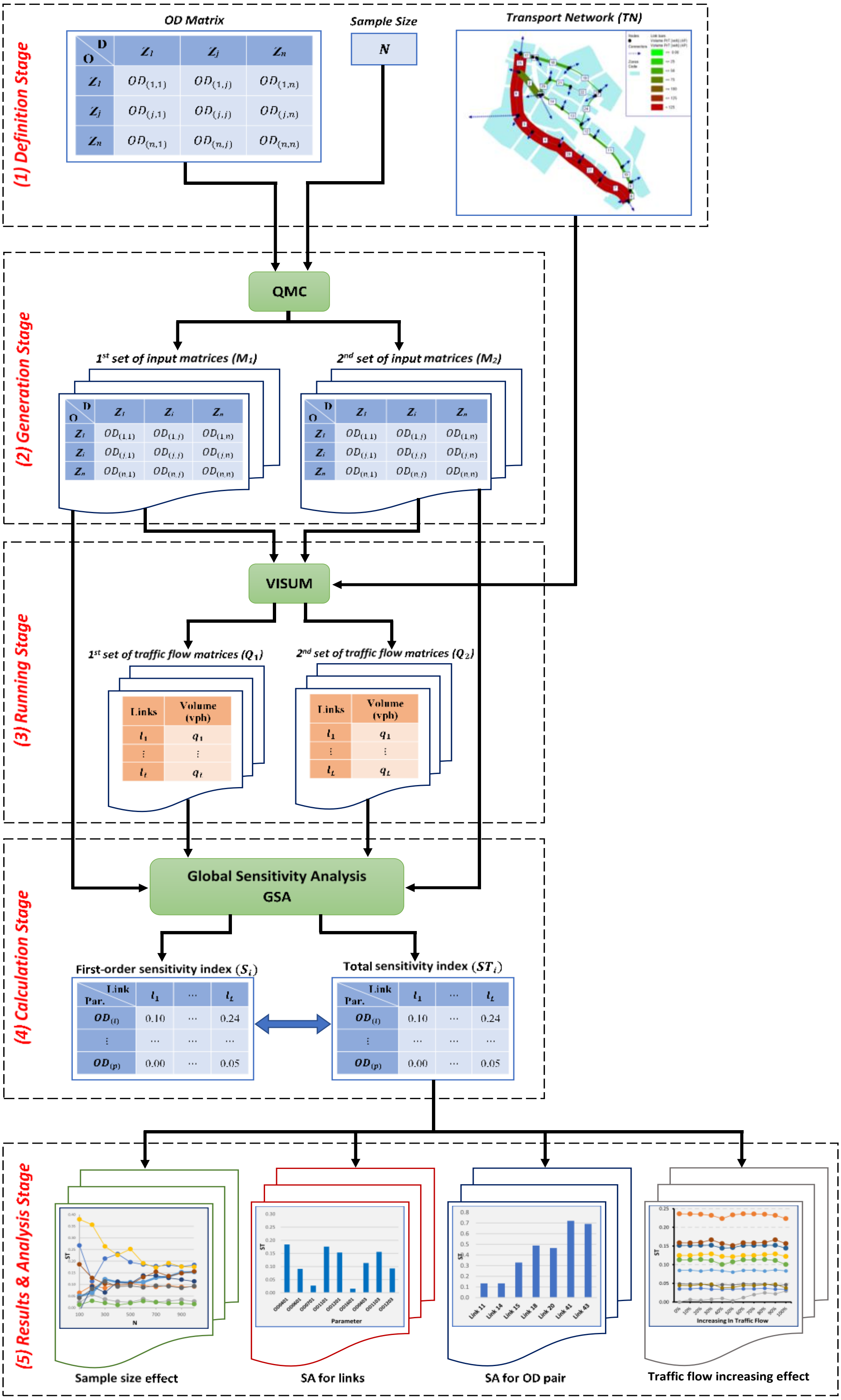


Figure 12: GSA-TAM algorithm of the traffic assignment model.

Algorithm 2. Sensitivity Analysis for Input Data of Traffic Assignment Model

Definition stage:

$Z_{j=1,n}$; $\forall Z \in \mathbb{N}$ ⊳ Zone definition.

$O_{j=1,n}$; $O \subset Z$ ⊳ Origin definition.

$D_{j=1,n}$; $D \subset Z$ ⊳ Destination definition.

$OD_{(j,j)}$; $\forall OD_{(j,j)} \in \mathbb{R}$ ⊳ Matrix of Origin Destination (OD).

$OD_{(j,j)}^{obs.}$; $\forall OD_{(j,j)}^{obs.} \in \mathbb{R}$ ⊳ Observed values of OD matrix.

$f_{T\,O\to D}$ ⊳ Traffic Assignment function.

$l = 1, L; \forall\, l, L \in \mathbb{N}$ ⊳ Links attribute.

L; ⊳ Number of links.

$N; N \in \mathbb{N}$ ⊳ Number of samples.

$p = Z^2$; $p \in \mathbb{N}$ ⊳ Number of parameters

Generation stage:

$\{\varphi_k^{M_1}\}^{N\times p} \leftarrow QMC$ [Sobol's sequence] ⊳ Generating samples for the 1st. input matrix (M_1).

$\{\varphi_k^{M_2}\}^{N\times p} \leftarrow QMC$ [Sobol's sequence] ⊳ Generating samples for the 2nd. Input matrix (M_2).

For $k = 1$ *to* N **do**

  For $i = 1$ *to* n **do**

    For $j = 1$ *to* n **do**

      $\{OD_{(j,j)}^k\}^{M_1} \leftarrow OD_{(j,j)}^{obs.} * \{\varphi_k^{M_1}\}$ ⊳ Get the 1st. set of input OD matrices $\{M_1\}$.

      $\{OD_{(j,j)}^k\}^{M_2} \leftarrow OD_{(j,j)}^{obs.} * \{\varphi_k^{M_2}\}$ ⊳ Get the 2nd. set of input OD matrix $\{M_2\}$.

    end for

  end for

end for

Running stage:

For $k = 1$ *to* N **do** ⊳ Run VISUM.

  $\{q_{l=1,L}^k\}^{M_1} \leftarrow f_{T\,O\to D}\{OD_{(j,j)}^k\}^{M_1}$; $\forall\, q_l^k \in \mathbb{R}$ ⊳ Get the 1st. set of traffic flow attributes $\{Q_1\}$.

  $\{q_{l=1,L}^k\}^{M_2} \leftarrow f_{T\,O\to D}\{OD_{(j,j)}^k\}^{M_2}$; $\forall\, q_l^k \in \mathbb{R}$ ⊳ Get the 2nd. set of traffic flow attributes $\{Q_2\}$.

end for

Calculation stage:

For $l = 1$ *to* L **do** ⊳ The first loop for links.

  For $i = 1$ *to* p **do** ⊳ The second loop for parameters.

    For $k = 1$ *to* N **do** ⊳ The third loop for samples.

      $\{OD_{(i=1:p)}^k\}^{V_1} \leftarrow reshape\ \{OD_{(i=1:n,j=1:n)}^k\}^{M_1}$ ⊳ Reshape matrix (M_1) to vector (V_1).

      $\{OD_{(i=1:p)}^k\}^{V_2} \leftarrow reshape\ \{OD_{(i=1:n,j=1:n)}^k\}^{M_2}$ ⊳ Reshape matrix (M_2) to vector (V_2).

      $\{OD_{(i=1:p)}^k\}^{V_3} \leftarrow \{OD_{(i=1:p)}^k\}^{V_2}$ ⊳ Generate vector (V_3).

      $\{OD_{(i)}^k\}^{V_3} \leftarrow \{OD_{(i)}^k\}^{V_1}$ ⊳ Replace the (i) column of (V_3) by the same column from (V_1).

      $\{OD_{(j=1:n,j=1:n)}^k\}^{M_3} \leftarrow reshape\ \{OD_{(i=1:p)}^k\}^{V_3}$ ⊳ Reshape vector (V_3) to matrix (M_3).

      $\{q_{l=1,L}^k\}^{M_3} \leftarrow f_{T\,O\to D}\{OD_{(i,j)}^k\}^{M_3}$; $\forall\, q_l^k \in \mathbb{R}$ ⊳ Run VISUM to get the 3rd. set of traffic flow attributes $\{Q_3\}$.

    end for

    $f_0^2 \leftarrow \left(\frac{1}{N}\sum_{k=1}^{N}\{q_l^k\}^{M_1}\right)^2$ ⊳ Constant.

    $S_i \leftarrow \dfrac{\left[\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(\{q_l^k\}^{M\,1}\right)^2 - f_0^2\right] - \left[\left(\frac{1}{2N}\right)\sum_{k=1}^{N}\left(\{q_l^k\}^{M\,2} - \{q_l^k\}^{M\,3}\right)^2\right]}{\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(\{q_l^k\}^{M\,1}\right)^2 - f_0^2}$ ⊳ First order sensitivity index.

    $ST_i \leftarrow \dfrac{\left(\frac{1}{2N}\right)\sum_{k=1}^{N}\left(\{q_l^k\}^{M\,1} - \{q_l^k\}^{M\,3}\right)^2}{\left(\frac{1}{N}\right)\sum_{k=1}^{N}\left(\{q_l^k\}^{M\,1}\right)^2 - f_0^2}$ ⊳ Total sensitivity index.

  end for

end for

End

3.4. Implementation

The proposed methodology has been applied to the case study (Ajka, Hungary). The following subsections represent the implementation stages:

3.4.1. Definition Stage

The methodology was implemented in the study case. The definition details of the study area are as follows:
Model: VISUM software / Traffic assignment model / Incremental assignment method.
Parameters (p): OD pairs, the number of parameters equal to the number of OD matrix cells; ($p = Z^2 = 625$ parameters).
The number of links: ($L = 50$).
The output result vector (Q), $Q = \{q_1, q_2, q_3, \dots, q_{50}\}$, represent traffic flow in the links.
Sample size (N): In this study, ten sizes of samples have been checked in purpose to get more reliable results. The samples ranged from 100 to 1000.

3.4.2. Generation Stage

Two matrices have been generated (M_1) and (M_2), each one has a dimension (N, p), by applying the QMC Sobol sequence. Because of 10 samples have been used in this case study, the dimensions of these matrices were {(100,625), (200,625), ..., (1000,625}. Thus, the total number of generated matrices to use in the present chapter are 20 matrices. Also, as noticeable as these matrices' dimensions are big, that directly leads to an increase in simulation time, representing the difficult challenge in this algorithm.

3.4.3. Running Stage

In this stage, The MATLAB code was executed twice multiplied by the number of samples for matrices (M_1) and (M_2) for all sampling cases. The results of this stage are two sets of output vectors: (Q_1) and (Q_2) corresponding to matrices (M_1) and (M_2). Each set has a dimension of ($p = 50$). This process was repeated ten times, according to the number of sample sizes (N) adopted from the previous stage.
The running stage consists of three steps: the first step is using the generated matrices (M_1) and (M_2) to forming in (2N=2000) of OD matrices. The second step is inputting 2000 OD matrices in VISUM to produce 2000 traffic flow attributes. The last step is forming

these attributes into two matrices (Q_1) and (Q_2), corresponding to the two sets of matrices (M_1) and (M_2) respectively. Each of (Q_1) and (Q_2) has a dimension (L, N).

3.4.4. Calculation Stage

This stage is the core of the GSA-TAM algorithm, where the first and total sensitivity indices are determined for all parameters by using Equation 25 and Equation 26. The most significant challenges in this stage: first, forming and changing the parameters of the matrix (M_3) for every sample during the simulation process. Second, the simulation time, although the computations were performed in the workstation computer has an Intel Core 6 CPU at 3.70 GHz with 32 GB RAM. The whole simulation process of GSA-TAM required more than 2000 hours to finish all samples size.

The results of this stage are two sets of output metrics: (S_i) for the first-order sensitivity index and (ST_i) for the total sensitivity index. Every set has a dimension of $(p = 625, L = 50)$. Finally, the calculation process was replicated ten times to determine the adequate sample size for this case study.

3.4.5. Analysis Stage

For the analysis stage, four questions will be answered. The first question is, what is the proper sample size for the sensitivity analysis of the traffic assignment model? Secondly, what are the OD pairs that affect each link? Thirdly, for every OD pair, what is the scope of effectiveness on the transport network? Finally, does the sensitivity indices change in case of increasing traffic flow? All these questions have been discussed in the following subsections.

3.4.5.1. Sample Size

Knowing and determining the proper sample size has a significant role in decreasing the GSA_TAM algorithm's required time. For this research, ten sizes of samples have been used to reach the minimum sample size. The minimum sample size is satisfying when the Sobol sensitivity indices have low variability between the previous and subsequent results. In this study case, the transport network has 50 links; only two links (l_{02}) and (l_{12}) will be presented to show the relationship between the sample size and Sobol total sensitivity index (see Figure 13 and Figure 14). In general, after all the 50^{th} links have been checked, the study found that the minimum size of samples required is $(N > 1000)$ for this case study.

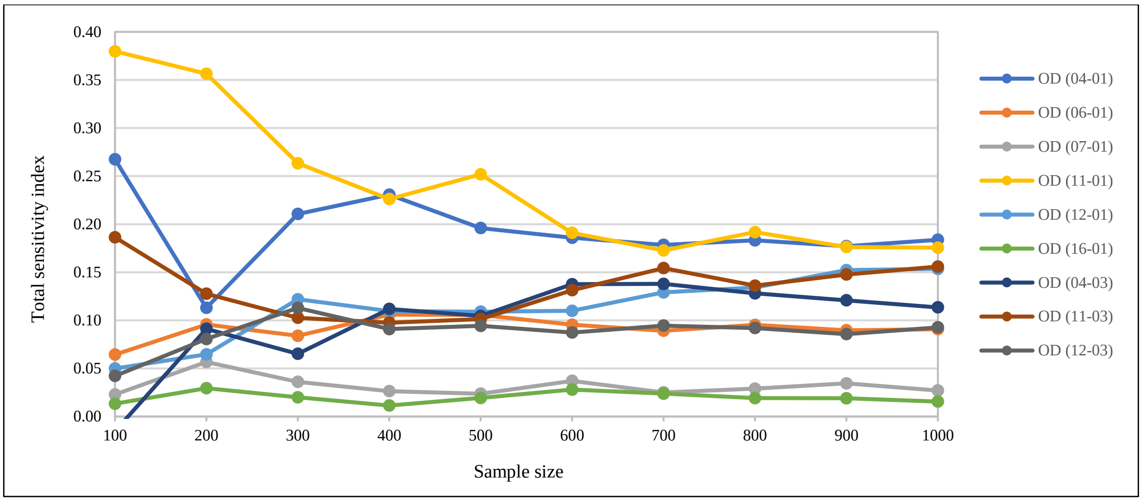


Figure 13: Evaluation of the total sensitivity index for (l_{02}).

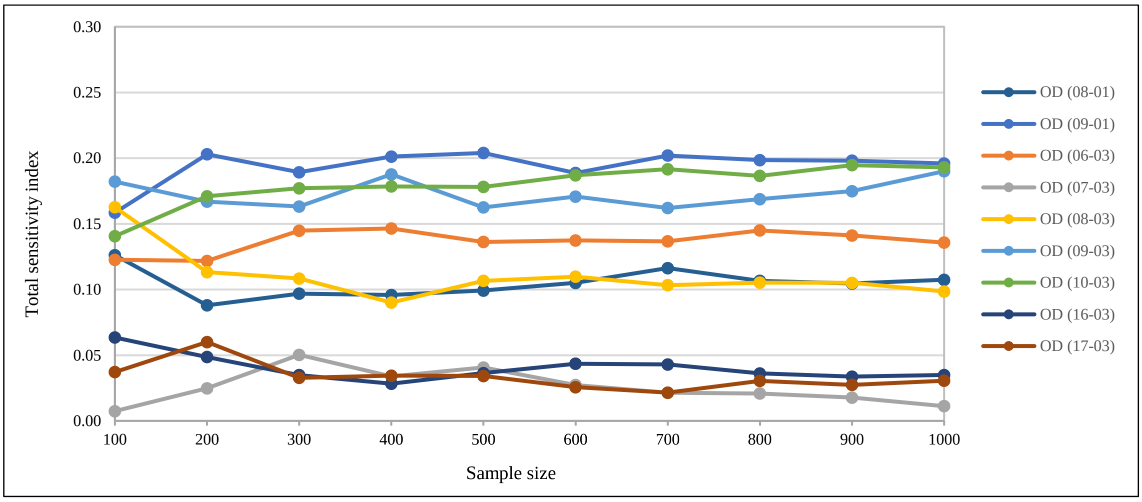


Figure 14: Evaluation of the total sensitivity index for (l_{12}).

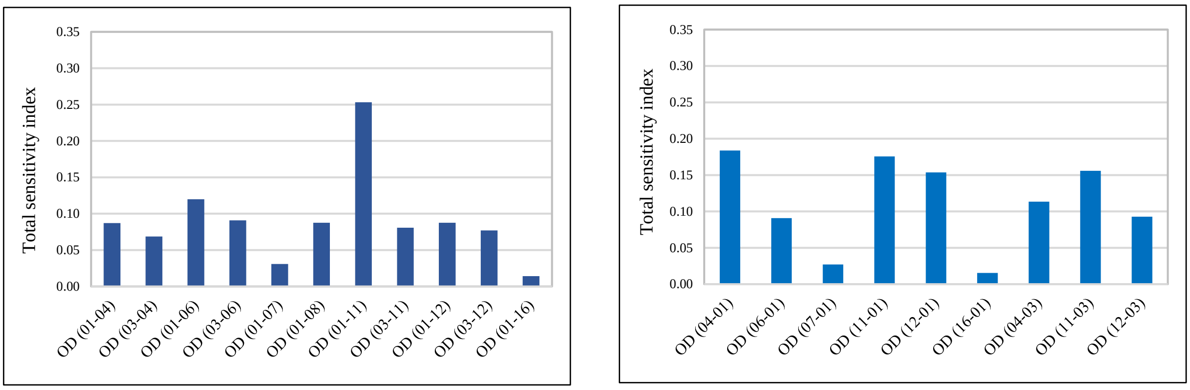
3.4.5.2. Total Sensitivity Index

This research aims to use the GSA-TSA algorithm to identify OD pairs' set that creates traffic flow in each transport network link. Besides, determining the influence ratio of each OD pair on traffic flow variance in any link.

Figure 15 showing the total sensitivity index (ST) of OD pairs in only four links out of 50 links forming the transport network of the study area by using the sample size ($N = 1000$). In this figure, the OD pairs that have the ($ST < 0.005$) are skipped showing.

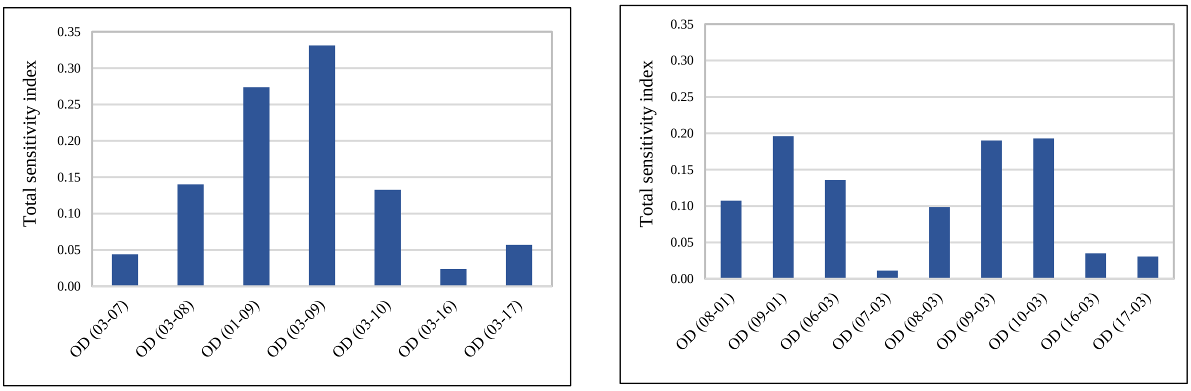
In general, each link has its own sets t of OD pairs affecting it. The number of OD pairs in each set is different from a link to another; according to the link's location in the transport network, the internal links have more OD pairs than the external links. For example, in Figure 15-a, the set of OD pairs affecting in the (l_{01}) consists of 11 OD pairs, While in

Figure 15-c, the set of OD pairs affecting in the (l_{11}) was 7 OD pairs only. Also, in Figure 15-a, it can be seen the (ST) indices for the parameters were variated from 0.253 for (OD_{01-11}) to 0.014 for (OD_{01-16}). As well as all parameters in other links.



(a) Total sensitivity index of OD pairs in (l_{01}).

(b) Total sensitivity index of OD pairs in (l_{02}).



(c) Total sensitivity index of OD pairs in (l_{11}).

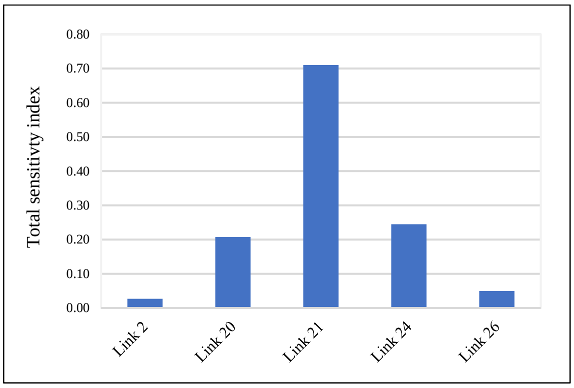
(d) Total sensitivity index of OD pairs in (l_{12}).

Figure 15: Total sensitivity index of OD pairs in several links on the transport network of the study case by using (N=1000): (a) Link # 01, (b) Link # 02, (c) Link # 11, and (d) Link # 12.

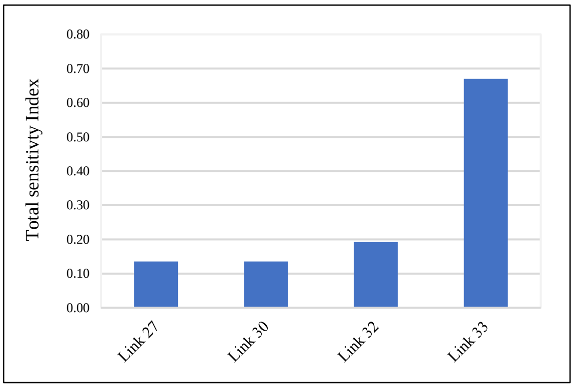
3.4.5.3. Parameter Scope

The other important goal of this research is to know the scope of effect of each OD pair on the whole links of the transport network. For this reason, the developed GSA algorithm for traffic assignment models included a technique to detect and define the affected domain by each OD pair.

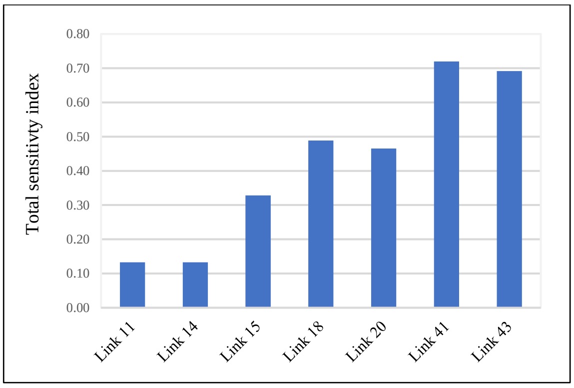
In this case study, the OD pairs' impact scope was ranged between (1 - 11) links. The impact scope of OD pairs is ruled mainly by the spatial distribution for traffic analysis zones (TAZ). If the OD zones are located close to each other, it decreases the required links to serve the traffic movement between them. Moreover, vice versa, if the OD zones are located far from each other. Figure 16 showing the impact scope for only four OD pairs out of 625 OD pairs in the study area by using the sample size ($N = 1000$). In Figure 16-a, the impact scope of OD (07-01) covering five links. And the (ST) Index of OD (07-01) differed from 0.027 in the (link # 02) to 0.710 in the (link # 21). And so on the other links.



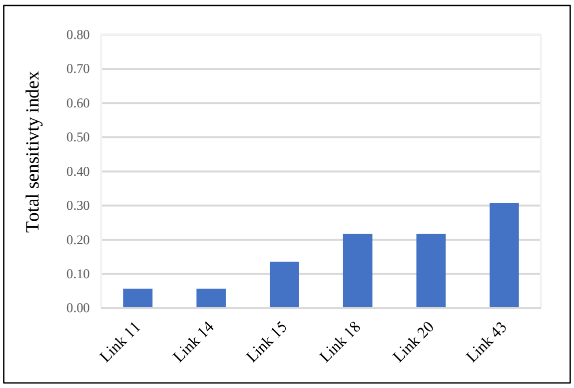
(a) The impact scope of (OD_{07-01}).



(b) The impact scope of (OD_{10-01}).



(c) The impact scope of (OD_{03-10}).



(d) The impact scope of (OD_{03-17}).

Figure 16: The impact scope of several OD pairs on the transport network of the study case by using ($N = 1000$): (a) OD_{07-01} (b) OD_{10-01} (c) OD_{03-10} and (d) OD_{03-17}.

3.4.5.4. Affectation of Increasing Traffic Flow

In this section, new simulations have been appropriated for the cases of an increase in traffic flow by increasing the travel demand frequently by 10% to reach 100% of the original travel demand (i.e., the base case). For each case, the GSA-TAM algorithm has been implemented. The simulation results show the effect of increasing traffic flow on the sensitivity indices of OD pairs.

In Figure 17, it can be seen that the sensitivity index of (OD_{11-03}) decreases from 0.210 in the base case to 0.147 in case of increasing travel demand up 100%; this means that a part of the travel demand from zone #11 to zone #03 is moved from (l_{02}) to another link(s). Furthermore, in Figure 18, it can be seen the reversed status, the sensitivity index of (OD_{01-03}) increases from zero (i.e., none effect) in the base case to 0.03 in case of increasing travel demand to 100%, and this means that a part of the travel demand from zone #01 to zone #03 is moved from other links to (l_{12}), and so on for the OD pairs.

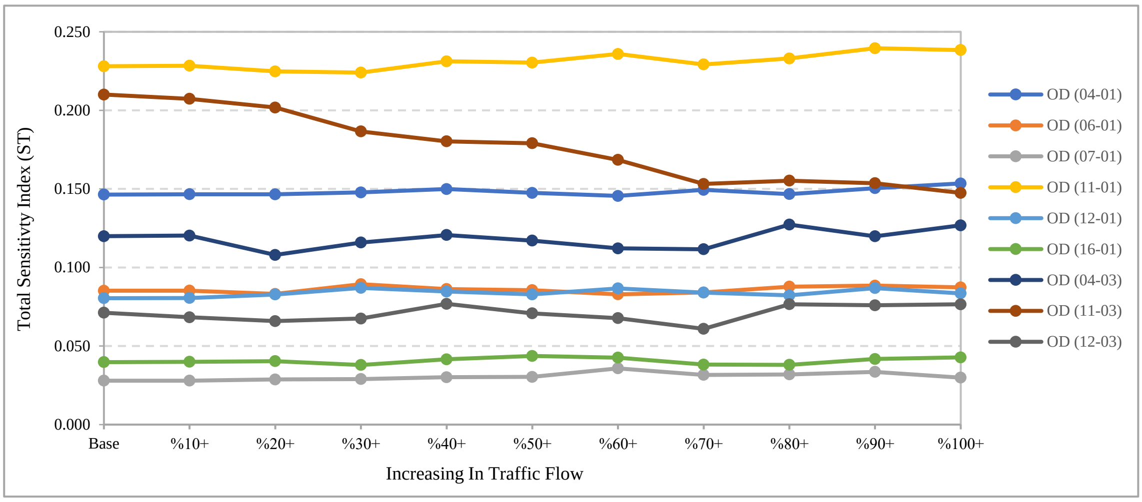


Figure 17: Evaluation of the total sensitivity indices for (l_{02}).

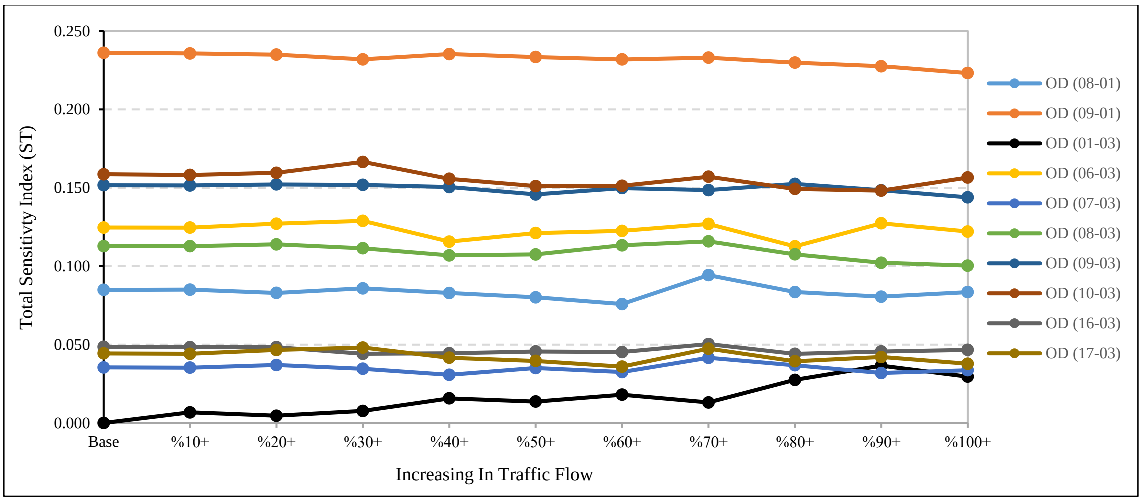


Figure 18: Evaluation of the total sensitivity indices for (l_{12}).

3.5. Conclusions

Global sensitivity analysis (GSA) quantifies the impact on the model output of a variation that occurred by model components; this allows recognizing the relationship between the changed component and the model output, for instance, importance, relevance, and effectiveness. In traffic assignment models, identifying the OD pairs that form traffic flow in each link is critical for quantifying traffic flow uncertainty causes and sources. Therefore, in the present chapter, a new methodology has been developed to obtain the OD pairs' sensitivity indices in all links of the transport network.

The main contribution of this chapter consists of presenting a new algorithm base on the variance-based method of global sensitivity analysis for the traffic assignment model (GSA-TAM algorithm). This algorithm is used to calculate Sobol sensitivity indices for the input OD matrix. The GSA-TAM algorithm has been written using MATLAB software and linked with VISUM software via the Component Object Model (COM). The GSA-TAM algorithm consisting of five following stages: (Definition Stage, Generation Stage, Running Stage, Calculation stage, and Analysis stage). A clear stepwise detail was afforded to implement each stage described in the algorithm.

The developed GSA-TAM algorithm aims to provide an applicable methodology for the following purposes:

- Detecting the set of OD pairs influencing on each link on the transport network.
- Identifying the effect value for each OD pair on the traffic flow in any link by knowing the sensitivity index.
- Classifying the OD pairs on each link according to sensitivity indices to show the more significant ones.
- Determining the practical scope for each OD pair on the transport network.

The significant challenge solved in this study was selecting a suitable sample size for this study. Therefore, ten samples have been used to examine OD pairs' sensitivity indices in the traffic assignment model. Depending on the sensitivity indices pattern for all samples, the minimum sample size meets with low variability between the previous and subsequent results is ($N > 1000$) for this case study.

4. Estimation of Link Choice Probabilities Using Maximum Likelihood Estimation Method and Gaussian Processes

4.1. Introduction

The fundamental premise in traffic assignment is choosing a route for each OD pair. Numerous factors influence a route choice; these include journey time, trip distance, monetary cost, congestion, type of roadway, characteristics, and perceptions of the drivers themselves. Containment of all these factors and the assumption that all drivers have full knowledge of the transport system is questionable. Accordingly, the finding of the route choice set for each traffic assignment model is a difficult task and not practical. Therefore approximations are inevitable in estimating the route choice set [62]. That approximation in the calculation of the route choice set caused uncertainty in the traffic assignment model on each link of the transport network, especially in links with higher traffic density [63, 64].

The drivers' route choice behaviour one of the main reasons causes traffic flows' fluctuations on transport networks. Therefore, it is essential to estimate the choice set in the traffic assignment model to forecast driver's behaviour under hypothetical scenarios, to predict future traffic conditions on transport networks, and to understand the uncertainty in the traffic assignment model. The importance of knowing the route choice set is increasingly recognized as a solution to uncertainty problems in the traffic assignment model. [65, 66].

Route choice sets are usually designated using choice set generation algorithms (e.g. shortest paths or labelling), which compute a set of routes based on properties of the transport network. These algorithms can carry two types of errors in the choice set generation. First, false-negative errors, which occur when the algorithm is not able to recreate the chosen alternatives. The created alternatives might not match the preferences and behaviour of the drivers, and as a result, the chosen route is not regenerate. The impact of this error decreases when the ability of choice set generation algorithm to capture the drivers' preferences and behaviour increases. Second, false-positive errors happen when a choice set generation algorithm also creates routes that are unconsidered by the divers, resulting in a too wide choice set. As a result, using choice set generation algorithms potentially comes with several imperfections [67].

Accordingly, the evaluation of the significance and reality of generated route choice sets is difficult in practice because the actual choice sets are unknown to the transport modellers. After all, it includes unlimited possible routes for each OD pair on the transport network. Several researchers, such as Ramming [68], Hoogendoorn-Lanser [69], Bekhor et al. [70], Hoogendoorn-Lanser and Nes [71], Bovy and Fiorenzo-Catalano [72] and Prato and Bekhor [73], have stated different methods to measure of quality and efficiency of the generated route choice sets. The empirical analysis shows that no choice set generation algorithm can fully reproduce observed paths.

Route choice sets are generally used to study for three application purposes: (i) Analysis of travel behaviour, where the planner is interested to know of travel alternatives, characteristics, number, variety, and composition. (ii) Estimating parameters of choice models (i.e., estimating parameters of utility functions of route choice models at an individual level). (iii) Predicting of choice probabilities for alternatives routes (i.e., determining route flow and link flow for travel demand analysis using route choice models) [72, 74].

To analysis uncertainty of the traffic flow required information to be gathered about what are the OD pairs using in each link (i.e., link choice set), and what are the probabilities proportion of the OD pairs that using this link (i.e., link choice probabilities). "Link choice" is closely related to "route choice" as the link is part of the route, and this study focuses on links instead of routes to analyze the characteristics of uncertainty.

For this study, the parameter estimation technique has been used for determining the most probable value of link choice probabilities for all links on the transport network. The parameter estimation technique plays a principal role in complex mathematical models, as it helps us to determine optimal parameters to create a reliable model [75]. Barnard and Bayes [76] described that for most mathematical models that have a large number of parameters, and the observed data was erroneous, Parameter estimation can be performed in the MLE framework, where the state estimate is the parameter which maximizes the likelihood function. The MLE framework gives estimates of the unknown quantities that maximize the probability of obtaining the observed set of data. MLE used in various scientific fields, such as communication systems, psychometrics, pharmacology, genetics, and astrophysics [77]. This work uses MLE to estimate choice probabilities sets for links by using a linear Gaussian model.

This chapter focuses on the analysis of link flow and its uncertainty characteristics. A new methodology has been innovated depending on MC simulation processes, to estimate of link choice probabilities for all links on the transport network, and using the methodology's results to find the error's parameters of traffic flow in the traffic assignment model. In this context, the new methodology based on both global sensitivity analysis of traffic assignment model algorithm GSA-TAM to find the generation set of OD pairs and maximum likelihood estimation MLE framework to estimate rote choice probabilities of these OD pairs in each link in the transport network.

4.2. Background

4.2.1. Route Choice Modelling

Choice set generation consists of finding all feasible routes that a traveller might consider for travelling from his origin to his destination. Choice set generation for route choice modelling is known as a complicated problem compared to other choice modelling problems such as mode choice or destination choice a good overview of route choice modelling can be found in Bovy and Stern [78], which is the first book entirely covering this topic.

It is well known that in a route choice context, finding route choice is complex for various reasons and involves several steps before the actual model estimation. The route choice model depends, on the one hand, on the attributes of the available routes, such as the type of road, speed limit, travel time, and the number of traffic lights. On the other hand, the characteristics and preferences of the divers also influence the choice. Some of the drivers like high speeds on freeways while others prefer small beautiful roads, some avoid left turns or traffic lights. Several aspects of the route choice problem make it particularly complicated [79, 80].

Over the years, various approaches have been suggested for modelling route choice behaviour, and these approaches can be subdivided into two categories deterministic and stochastic; Figure 19 illustrates the situation [79]. In this chart, U represents the universal set that includes all possible routes for an OD pair. In the deterministic approach, M represents the master set of known routes generated by the researcher using deterministic or stochastic methods in order to approximate a driver's preferences and awareness set. And then, if the choice set generation method is deterministic, the probability of a driver choosing route r simply equals $P(r|C)$ where C is the individual final viable choice set, where $(C \subseteq M)$. On the other hand, if the choice set generation method is a probabilistic approach where all non-empty subsets G are considered, G is all the non-empty subsets of the master set M. The probability probabilistic approach would be as follow:

$$P(r) = \sum_{C \in G} P(r|C)P(C) \tag{28}$$

Where:

$P(r)$	the probability of a driver choosing route r from master set M,
$P(r\|C)$	the probability of a driver is choosing route r from a given choice set C,
$P(C)$	probability of a driver's choice set being C.

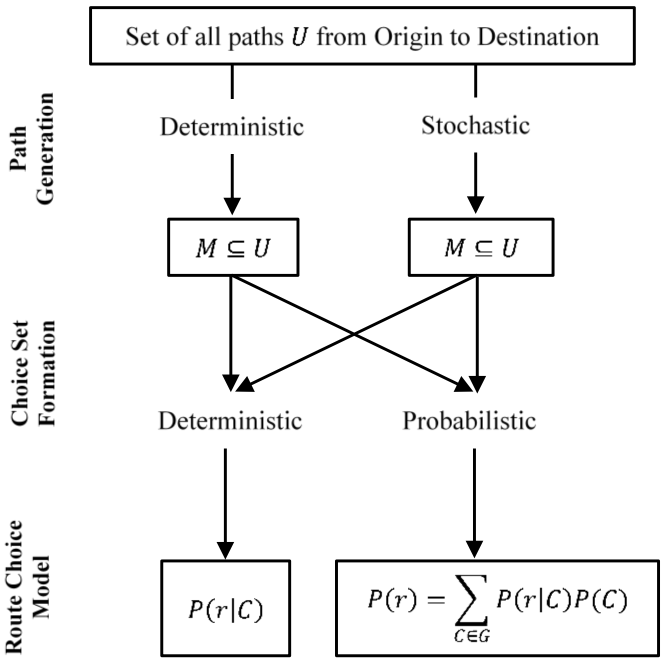


Figure 19: Choice set generation [79]

As noted above, the primary purpose of route choice models is predicting probabilities of alternatives routes/links. However, the generation of route choice sets suffers from several problems; First, it is unclear which criteria should be adopted to determine the appropriate alternatives, like the choice set, should have sufficient variety and include alternatives that are logically relevant for the study. Second, the route choice models usually consider trips between origin zones to destination zones, while estimation usually concentrates on individual trips. As a result, choice sets utilized for prediction should be much more significant to account for the differences in spatial and individual characteristics [67, 68, 74].

4.2.2. Maximum Likelihood Estimation

MLE is one of a widely common technique used in statistical, engineering, economics, and many different applications to estimate the parameters of a model [81]. MLE can be implemented for most mathematical models regardless of the number of parameters in these models. The MLE framework gives values to the parameters that maximize the likelihood function [77].

Generally, the MLE method estimates set of parameters, that give simultaneous solution to the set of equations, when the derivatives of the log-likelihood equations are set at zero. After that, the MLE technique involves obtaining all second derivatives of the likelihood for each parameter by employing any practical method of solution [77, 81, 82]. For a more precise description, the model output (Y) is a random vector which have a number of independently observations (N). Also, assume for now that (Y) is continuous, denote by $f(Y|\theta)$ the probability density function $pdf(Y)$, conditioned on a set of parameters (θ). The random distribution would be wholly characterized if the particular value of (θ) is known, say (θ_0), corresponding to the population under interest. Since the observations are assumed to be independent, the joint density is the product of the individual densities:

$$f(y_1, y_2, \dots, y_N|\theta) = \prod_{i=1}^{N} f(y_n|\theta) \tag{29}$$

Considering a function of θ that would follow the value of the joint density, given the observations $y_1, y_2, \dots, y_N$. The likelihood function becomes:

$$L(\theta|y_1, y_2, \dots, y_N) = f(y_1, y_2, \dots, y_N|\theta) \tag{30}$$

Since the (θ_0) is unknown; the solution can be done by approximating it through computing an estimator $(\hat{\theta}_N)$ of its value that can be ruled as the most likely value for θ, given the observations. The estimation is done by maximizing the function $L(\theta|y_1, y_2, \dots, y_N)$ concerning (θ):

$$\max_{\theta \in \Theta} L(\theta|y_1, y_2, \dots, y_N) \tag{31}$$

Where: Θ is parameter space, and the θ belongs to Θ.

Furthermore, that Eq. (11) has a unique solution, called the maximum likelihood estimator:

$$\hat{\theta}_N = \arg\max_{\theta \in \Theta} L(\theta|y_1, y_2, \dots, y_N) \tag{32}$$

practically, it is often more beneficial to solve with the logarithm of the likelihood function, which is called the log-likelihood:

$$\ln L(\theta|y_1, y_2, \dots, y_N) = \ln f(y_1, y_2, \dots, y_N|\theta) = \sum_{n=1}^{N} \ln f(y_i|\theta) \tag{33}$$

The maximum likelihood estimator under normal conditions, $\hat{\theta}_N$ almost definitely converges to θ_0 as N increases to infinity, and the distribution function of $\sqrt{N}$ $(\hat{\theta}_N - \theta_0)$ converges to the multi-normal distribution function with mean equal zero [81].

4.2.3. Gaussian Processes

The Gaussian processes GP has been commonly used in statistics and machine-learning studies for modelling stochastic processes in regression and classification [83]. The advantage of using the GP originates principally from two fundamental properties. First, a GP is wholly determined by its mean and covariance functions. This property promotes model fitting. Second, solving the prediction problem is almost straight- forward. The best predictor of a GP is a linear regression function of the observed values, and, in many cases, these functions can be computed by applying recursive formulas [84, 85].

GP provides a technique for modelling probability distributions over functions based on the framework of Bayesian regression [83, 86]. Formally, GP generates data located throughout some domain such that any finite subset of the range follows a Normal multivariate distribution. Consider the Bayesian regression model the standard linear regression model with Gaussian noise:

$$Y = f(X) + \varepsilon \tag{34}$$

where X is the input matrix, $f(X)$ is the function value, Y is the model outputs vector and ε is noise distributed independently and identically according to Normal distribution with zero mean value and variance σ^2, i.e.:

$$\varepsilon \sim N(0,\ \sigma^2) \tag{35}$$

For each observation, y can be thought of as related to an underlying function $f(x)$ through a Gaussian noise model:

$$Y = f(X) + N(0,\ \sigma^2) \tag{36}$$

The basic linear regression formula with Gaussian noise:

$$Y = \beta X + N(0,\ \sigma^2) \tag{37}$$

Where:

β the unknown deterministic parameter.

Hence, the output Y will inherit the Gaussian distribution from the noise term ε resulting in:

$$p(Y|\beta) = p_{\varepsilon}(Y - \beta X) = N\ (Y|\beta X,\ \sigma^2) \tag{38}$$

4.3. Methodology

In the literature of traffic assignment, all route choice models are looking for answers about the following two questions. The first question is what the possible routes for each OD pair are? Moreover, the second is what the probabilities of these routes are? However, in the new methodology, this chapter starts from a further two different questions. The first is what the OD pairs that pass by a link are? Furthermore, the second is what the probabilities of these OD pairs are? The peculiarity of this methodology it is starting from the link(s) to find the other variables (i.e., OD pairs with their probabilities). Therefore, this methodology enables us to discover the characteristics of traffic flow variation of links in the case of fluctuation in travel demand.

The following methodology consists of three stages: Firstly, collecting data about the travel demand of the study area. Secondly, generating data using the MC simulation process. Finally, parameter estimation to find the probabilities of the OD set in every link on the transport network.

The relationships connecting these stages of the methodology are presented in Figure 20. Also, all of the mathematical and logical computations of this methodology are illustrated in Algorithm 3.

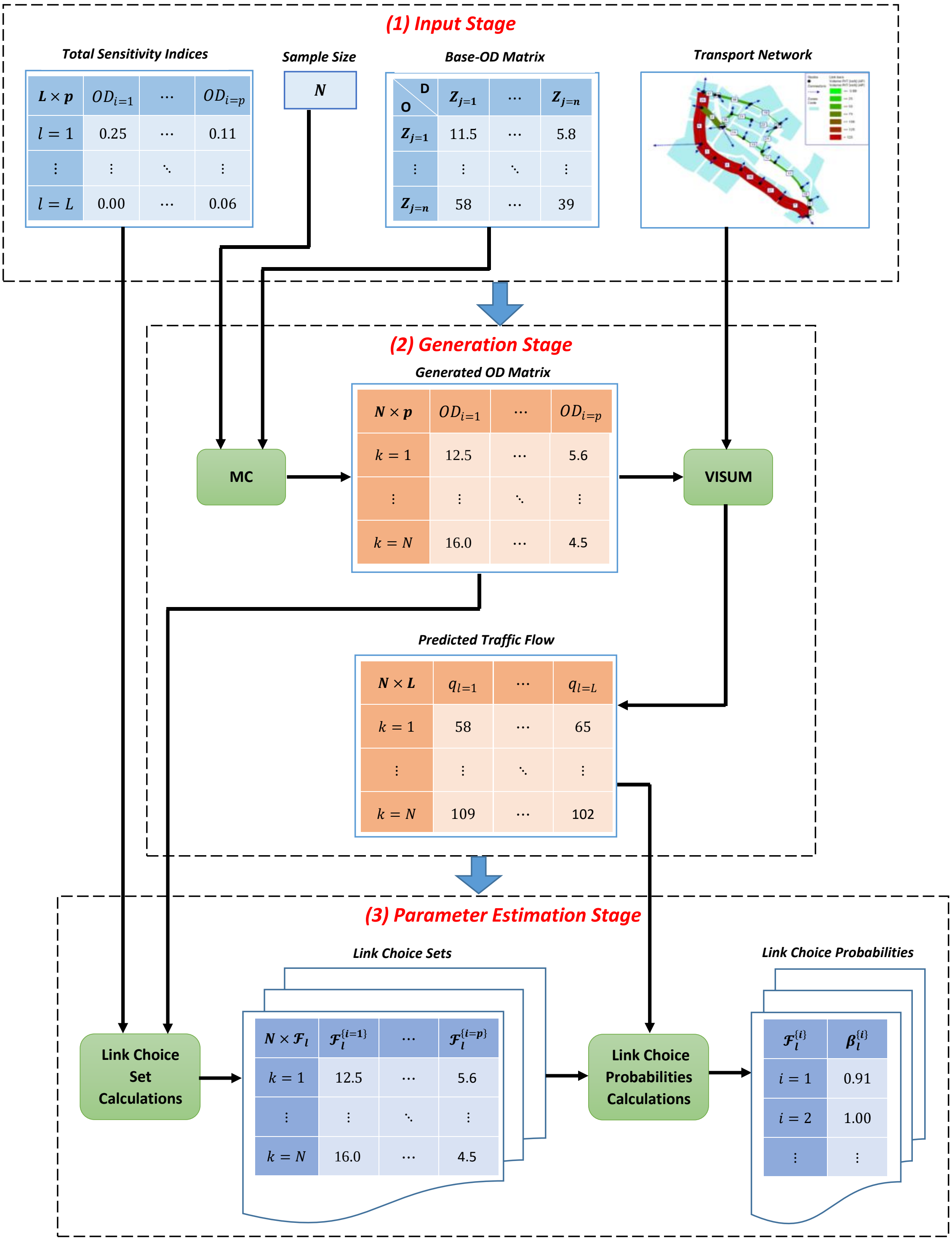


Figure 20: Flowchart illustrates the methodology of estimation of the link choice probabilities.

Algorithm 3. Parameter estimation of link choice probabilities

Input Data stage:

$Z_{j=1,n}$; $\forall\, Z, i \in \mathbb{N}$ ⊳ Zone definition.
$O_{j=1,n}$; $O \subset Z$ ⊳ Origin definition.
$D_{j=1,n}$; $D \subset Z$ ⊳ Destination definition.
$OD_{(n,n)}$; $\forall OD_{(i,j)} \in \mathbb{R}$ ⊳ Matrix of Origin Destination (OD).
$OD^{base}_{(n,n)}$; $\forall OD^{base}_{(j,j)} \in \mathbb{R}$ ⊳ base value of OD pair.
$f_{T\,O\to D}$ ⊳ Traffic Assignment function.
$L; \forall\, L \in \mathbb{N}$ ⊳ Number of links.
$p = Z^2;\ p \in \mathbb{N}$ ⊳ Number of OD pairs.
PD ⊳ Define the Probability Distribution type.
$\mu, \sigma;\ \forall\, \mu, \sigma \in \mathbb{R}$ ⊳ Parameters range for the probability distribution.
$N; N \in \mathbb{N}$ ⊳ Number of MC samples.

Generation stage:

$\{\varphi_k\} \leftarrow MC\ [PD\ (\mu, \sigma)]$ ⊳ Get samples by the defined probability distribution and parameters using Monte Carlo procedure.
For $k = 1\ to\ N$ do
  For $j = 1\ to\ n$ do
    For $j = 1\ to\ n$ do
      $OD^{k}_{(j,j)} \leftarrow OD^{base}_{(j,j)} * \{\varphi_k\}$ ⊳ Get the generated-OD matrices.
    end for
  end for
end for
For $k = 1\ to\ N$ do ⊳ Run Visum.
  $q^{k}_{l=1,L} \leftarrow f_{T\,O\to D}\left(OD^{k}_{(n,n)}\right);\ \forall\, q^{k}_{l} \in \mathbb{R}$ ⊳ Get the predicted traffic flow attribute.
end for

Parameter estimation stage:

For $l = 1\ to\ L$ **do** ⊳ The first loop for links.
  For $i = 1\ to\ p$ **do** ⊳ The second loop for OD pairs.
    If $ST_{(l,i)} \neq 0$
      $\mathcal{F}^{\{i\}}_{l} = OD^{\{i\}}$ ⊳ Link Choice set.
    end if
    $\beta^{\{i\}}_{l} = \left(\left(\mathcal{F}^{\{i,k=1,N\}}_{l}\right)^{T}\left(\mathcal{F}^{\{i,k=1,N\}}_{l}\right)\right)^{-1}\left(\mathcal{F}^{\{i,k=1,N\}}_{l}\right)^{T}\left(q^{(k=1,N)}_{l}\right)$ ⊳ Link choice probabilities.
  end for
end for
End

4.3.1. Data Stage

The required data for this stage consists of three components: the first is setting the transport network (TN) and traffic analysis zones (TAZ) in VISUM software. The second is defining the OD matrix. The last is specifying the sample size for the simulation process. The notations used in this chapter:

OD^{base}	base OD matrix,
p	the total number of OD pairs,
L	the total number of links,

N	sample size.
TSI	total sensitivity indices matrix,
PD	probability distribution for the Monte Carlo procedure,
RSD, σ	parameters range for the Monte Carlo procedure.

4.3.2. Generation Stage

For this methodology, a MATLAB code has been written and linked with the Component Object Model (COM) of VISUM software. Hence, the simulation code in this stage involves two processes: the first is the generation of the OD matrix ($OD^{k=1,N}$) by Monte Carlo method for a given type of probability distribution 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 road network from traffic assignment function.

4.3.3. Parameter estimation

In the last stage of this methodology, the route choice probabilities of each set $\mathcal{F}_l$ will be estimated using Gaussian Process GP based on the Maximum Likelihood ML framework. The traffic flow in any link l caused by the travel demand movement of a set of OD pairs and the probabilities set of these OD pairs that used this link l. Consider the Bayesian regression model the standard linear regression model with Gaussian noise; the traffic flow q_l for a link l:

$$q_l = \mathcal{F}_l \, \beta_l + \varepsilon_l \tag{39}$$

Where: q_l is the traffic flow in the link l, $\mathcal{F}_l$ is the link choice set of OD pairs for the link l, β_l is the link choice probabilities set for the link l, ε_l is an error follows the normal distribution with a zero mean value and variance ${\sigma^2}_l$, hence $\varepsilon_l \sim N(0, \, {\sigma^2}_l)$. Thus, the normal density function for the ε_l:

$$N(\varepsilon_l | o, \, {\sigma^2}_l) = \frac{1}{\sqrt{2\pi \, {\sigma^2}_l}} exp\left(-\frac{1}{2\sigma^2}(q_l - \, \mathcal{F}_l \, \beta_l)^2\right) \tag{40}$$

The likelihood function is a joint density of $\prod_{k=1}^{N} f({\varepsilon_l}^k)$. Therefore, the likelihood function:

$$L(\beta_l, \, {\sigma^2}_l | q_l, \mathcal{F}_l) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi \, {\sigma^2}_l}} exp\left(-\frac{1}{2\,{\sigma^2}_l}\left({q_l}^k - \, {\mathcal{F}_l}^k \, \beta_l\right)^2\right) \tag{41}$$

The ML framework will obtain the estimates of β_l and ${\sigma^2}_l$ by maximizing the likelihood function.

First, the logarithm of this function:

$$\ln L(\beta_l,\ \sigma^2{}_l) = -\frac{N}{2}\ln(2\pi) - N\ln(\ \sigma^2{}_l) - \frac{1}{2\ \sigma^2{}_l}(q_l - \mathcal{F}_l\beta_l)^T(q_l - \mathcal{F}_l\beta_l) \quad (42)$$

Second, derivatives the Equation 42 concerning β_l and $\sigma^2{}_l$ parameters $\frac{\partial L}{\beta_l} = 0\ and\ \frac{\partial L}{\partial\ \sigma^2{}_l} = 0$ respectively.

The estimated probabilities set $\widehat{\beta_l}$:

$$\widehat{\beta_l} = (\mathcal{F}_l{}^T\mathcal{F}_l)^{-1}\ \mathcal{F}_l{}^T q_l \quad (43)$$

The estimated variance $\widehat{\sigma^2}_l$:

$$\widehat{\sigma^2}_l = \frac{1}{N}(q_l - \mathcal{F}_l\widehat{\beta_l})^T(q_l - \mathcal{F}_l\widehat{\beta_l}) \quad (44)$$

The predicted traffic flow $\widehat{q_l}$.

$$\widehat{q_l} = \mathcal{F}_l\ \widehat{\beta_l} \quad (45)$$

The estimated error $\widehat{\varepsilon_l}$:

$$\widehat{\varepsilon_l} = q_l - \widehat{q_l} \quad (46)$$

4.4. Implementation

4.4.1. Synthetic example

To full-understanding of the present methodology, a synthetic example for a small network consists of 4 zones and 22 links, as shown in Figure 21. The links' length is proportional to the measurement in the figure, the movement is allowed in all directions at all nodes, and all the links are two-lane two-way. According to the methodology, estimating link choice probabilities for link #15 required the following stages.

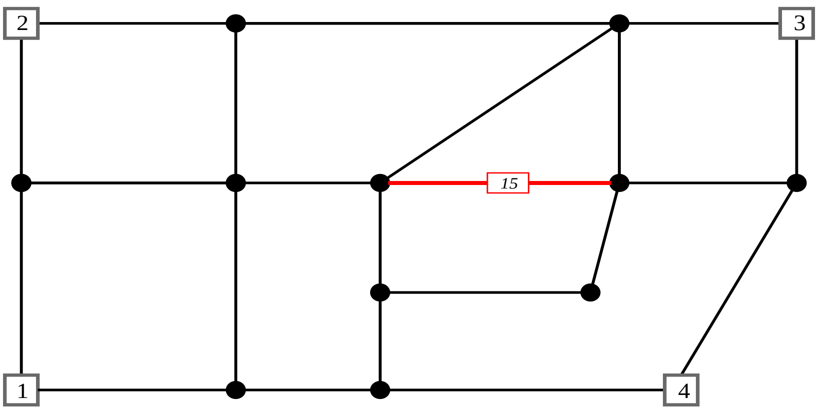


Figure 21: Synthetic example network.

Data stage: for this stage, a definition is required to OD matrix as shown in Table 3, the total sensitivity indices (ST) for OD pairs on link #15, as shown in Table 4, the sample size (assume N=1000), and the transport network definition in VISUM software.

Table 3: OD matrix (vph).

	Z_1	Z_2	Z_3	Z_4
Z_1	0.00	250	400	250
Z_2	200	0.00	250	200
Z_3	150	200	0.00	150
Z_4	300	250	200	0.00

Table 4: Total sensitivity indices of OD pairs for link # 15.

	Z_1	Z_2	Z_3	Z_4
Z_1	0.00	0.00	0.60	0.00
Z_2	0.00	0.00	0.07	0.18
Z_3	0.10	0.00	0.00	0.00
Z_4	0.00	0.05	0.00	0.00

Generation stage: the MATLAB code has been applied to generate input matrices sets (M) and executing these matrices sets via VISUM software to produce traffic flow attributes sets (Q).

$$M = \begin{bmatrix} OD_{1-1}^{1} & OD_{1-2}^{1} & OD_{1-3}^{1} & \cdots & OD_{4-4}^{1} \\ OD_{1-1}^{2} & OD_{1-2}^{2} & OD_{1-3}^{2} & \cdots & OD_{4-4}^{2} \\ OD_{1-1}^{3} & OD_{1-2}^{3} & OD_{1-3}^{3} & \cdots & OD_{4-4}^{3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ OD_{1-1}^{1000} & OD_{1-2}^{1000} & OD_{1-3}^{1000} & \cdots & OD_{4-4}^{1000} \end{bmatrix}$$

$$Q = \begin{bmatrix} q_1^1 & q_2^1 & q_3^1 & \cdots & q_{22}^1 \\ q_1^2 & q_2^2 & q_3^2 & \cdots & q_{22}^2 \\ q_1^3 & q_2^3 & q_3^3 & \cdots & q_{22}^3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_1^{1000} & q_2^{1000} & q_3^{1000} & \cdots & q_{22}^{1000} \end{bmatrix}$$

Parameter estimation stage: in this stage, the total sensitivity indices (TSI) matrix were used to establish the link choice set ($\mathcal{F}_{15}$) by ditching from the OD pairs that have zero sensitivity indices. The link choice set ($\mathcal{F}_{15}$) consists of 5 OD pairs, as follow:

$\mathcal{F}_{15} = \{OD_{1-3}, OD_{2-3}, OD_{2-4}, OD_{3-1}, OD_{4-2}\}$

The link choice set ($\mathcal{F}_{15}$) for 1000 samples, as follow:

$$\mathcal{F}_{15} = \begin{bmatrix} OD_{1-3}^{1} & OD_{2-3}^{1} & OD_{2-4}^{1} & OD_{3-1}^{1} & OD_{4-4}^{1} \\ OD_{1-3}^{2} & OD_{2-3}^{2} & OD_{12-4}^{2} & OD_{3-1}^{2} & OD_{4-4}^{2} \\ OD_{1-3}^{3} & OD_{2-3}^{3} & OD_{2-4}^{3} & OD_{3-1}^{2} & OD_{4-4}^{3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ OD_{1-3}^{1000} & OD_{2-3}^{1000} & OD_{2-4}^{1000} & OD_{3-1}^{1000} & OD_{4-4}^{1000} \end{bmatrix}$$

The calculated traffic flow (q_{15}) for 1000 samples, as follow:

$$\mathrm{q}_{15} = \begin{bmatrix} q_{15}^{1} \\ q_{15}^{2} \\ q_{15}^{3} \\ \vdots \\ q_{15}^{1000} \end{bmatrix}$$

So, the link choice probabilities set (β_{15}) has been estimated by applying Equation 43.

$$\begin{bmatrix} \beta_{1-3} \\ \beta_{2-3} \\ \beta_{2-4} \\ \beta_{3-1} \\ \beta_{4-2} \end{bmatrix} = \left[\begin{bmatrix} OD_{1-3}^{1} & OD_{2-3}^{1} & OD_{2-4}^{1} & OD_{3-1}^{1} & OD_{4-4}^{1} \\ OD_{1-3}^{2} & OD_{2-3}^{2} & OD_{12-4}^{2} & OD_{3-1}^{2} & OD_{4-4}^{2} \\ OD_{1-3}^{3} & OD_{2-3}^{3} & OD_{2-4}^{3} & OD_{3-1}^{2} & OD_{4-4}^{3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ OD_{1-3}^{1000} & OD_{2-3}^{1000} & OD_{2-4}^{1000} & OD_{3-1}^{1000} & OD_{4-4}^{1000} \end{bmatrix}^{T} \begin{bmatrix} OD_{1-3}^{1} & OD_{2-3}^{1} & OD_{2-4}^{1} & OD_{3-1}^{1} & OD_{4-4}^{1} \\ OD_{1-3}^{2} & OD_{2-3}^{2} & OD_{12-4}^{2} & OD_{3-1}^{2} & OD_{4-4}^{2} \\ OD_{1-3}^{3} & OD_{2-3}^{3} & OD_{2-4}^{3} & OD_{3-1}^{2} & OD_{4-4}^{3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ OD_{1-3}^{1000} & OD_{2-3}^{1000} & OD_{2-4}^{1000} & OD_{3-1}^{1000} & OD_{4-4}^{1000} \end{bmatrix} \right]^{-1}$$

$$\begin{bmatrix} OD_{1-3}^{1} & OD_{2-3}^{1} & OD_{2-4}^{1} & OD_{3-1}^{1} & OD_{4-4}^{1} \\ OD_{1-3}^{2} & OD_{2-3}^{2} & OD_{12-4}^{2} & OD_{3-1}^{2} & OD_{4-4}^{2} \\ OD_{1-3}^{3} & OD_{2-3}^{3} & OD_{2-4}^{3} & OD_{3-1}^{2} & OD_{4-4}^{3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ OD_{1-3}^{1000} & OD_{2-3}^{1000} & OD_{2-4}^{1000} & OD_{3-1}^{1000} & OD_{4-4}^{1000} \end{bmatrix}^{T} \begin{bmatrix} q_{15}^{1} \\ q_{15}^{2} \\ q_{15}^{3} \\ \vdots \\ q_{15}^{1000} \end{bmatrix}$$

The result of this synthetic example is presented in Table 5. In this synthetic example, the traffic flow in link #15 produced by proportions (β_{15}) of travel demand of a set of OD pairs $\mathcal{F}_{15}$. Practically, from the results in (Table 5), the estimated traffic flow ($\hat{q}_{15}$) can be calculated by applying Equation 45.

Table 5: The link choice set and link choice probabilities for link # 15.

Link choice set $\mathcal{F}_{15}$	OD_{1-3}	OD_{2-3}	OD_{2-4}	OD_{3-1}	OD_{4-2}
Traffic demand (vph)	100	250	200	150	150
Link choice probabilities set $\{\beta\}$	β_{1-3}	β_{2-3}	β_{2-4}	β_{3-1}	β_{4-2}
Link choice probabilities	0.38	0.04	0.20	0.13	0.03
Traffic flow per each OD pair (vph)	38	10	40	19.5	4.5
Predicted traffic flow $\hat{q}_{15}$ (vph)			112		

4.4.2. Case Study

The proposed methodology has been applied to the case study (Ajka, Hungary). The following subsections represent the implementation stages:

4.4.2.1. Data Stage

The required data for this stage includes the following:

- The base-OD matrix (OD^{base}),
- The total sensitivity indices (TSI),
- In this study, ten scenarios have been examined, the MC parameters:
  - PD: Normal distribution,
  - $\mu = 1$,
  - $RSD = (0.05, 0.10, \dots, 0.50)$,
- Sample size: (N=1000),
- Transport network definition in VISUM software.

4.4.2.2. Generation Stage

The MATLAB code has been applied ten times to generate input matrices sets ($M_1, M_1, \dots, M_{10}$) corresponding to $RSD = (0.05, 0.10, \dots, 0.50)$. Next, executing these matrices sets via VISUM software to produce traffic flow attributes sets ($Q_1, Q_1, \dots, Q_{10}$) respectively.

4.4.2.3. Parameter Estimation

The study has sufficed to show the analysis results for two links only; the selected links are (l_{02} and l_{12}). These links summarize the traffic flow pattern of other links in the study area. Moreover, ten scenarios have been tested to evaluate the effect of variation on the link choice sets ($\mathcal{F}_{02}$ and $\mathcal{F}_{12}$) and link choice probabilities set (β_{02} and β_{12}) for the selected links.

Table 6 and Table 7 illustrating the link choice set ($\mathcal{F}_{02}$ and $\mathcal{F}_{12}$) and the values of link choice probabilities (β_{02} and β_{12}) in all scenarios. While Figure 22 and Figure 23 show the sequential graphs: (1) The calculated traffic flow by the proposed methodology via VISUM software, (2) The predicted traffic flow by applying the model in Equation 45, (3) The cross-validation relationship between the calculated traffic flow via VISUM and the predicted traffic flow via the model, (4) The predicted error by applying Equation 46.

Table 6: Link choice set and link choice proportions set of the link β_{02} for various RSD.

$\mathcal{F}_{02}$	β_{02}	RSD									
		0.05	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
		Link choice proportions values									
OD_{04-01}	β_{04-01}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9931	0.9849
OD_{05-01}	β_{05-01}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9930	0.9763	0.9881	0.9587	1.0000
OD_{06-01}	β_{06-01}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	1.0000	1.0000	1.0000	0.9807
OD_{07-01}	β_{07-01}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9970	1.0000
OD_{11-01}	β_{11-01}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9948	0.9974	0.9917	0.9783
OD_{12-01}	β_{12-01}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9967	1.0000	1.0000
OD_{16-01}	β_{16-01}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9973	0.9960	1.0000	1.0000
OD_{04-02}	β_{04-02}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9971	1.0000	1.0000	1.0000	1.0000
OD_{05-02}	β_{05-02}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9757	1.0000	1.0000	1.0000	1.0000
OD_{06-02}	β_{06-02}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9986	1.0000	1.0000	1.0000
OD_{07-02}	β_{07-02}	1.0000	1.0000	1.0000	1.0000	0.9997	1.0000	0.9782	1.0000	1.0000	0.9644
OD_{11-02}	β_{11-02}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9968	0.9954	1.0000	1.0000
OD_{12-02}	β_{12-02}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988	1.0000	1.0000	1.0000	1.0000
OD_{16-02}	β_{16-02}	1.0000	1.0000	1.0000	1.0000	0.9999	0.9989	1.0000	1.0000	1.0000	1.0000
OD_{17-02}	β_{17-02}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
OD_{04-03}	β_{04-03}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9945	0.9998	0.9811	0.9968
OD_{05-03}	β_{05-03}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9956	0.9881	1.0000	1.0000	1.0000
OD_{11-03}	β_{11-03}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9975	0.9939	0.9895	0.9853
OD_{12-03}	β_{12-03}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9969	1.0000	0.9873

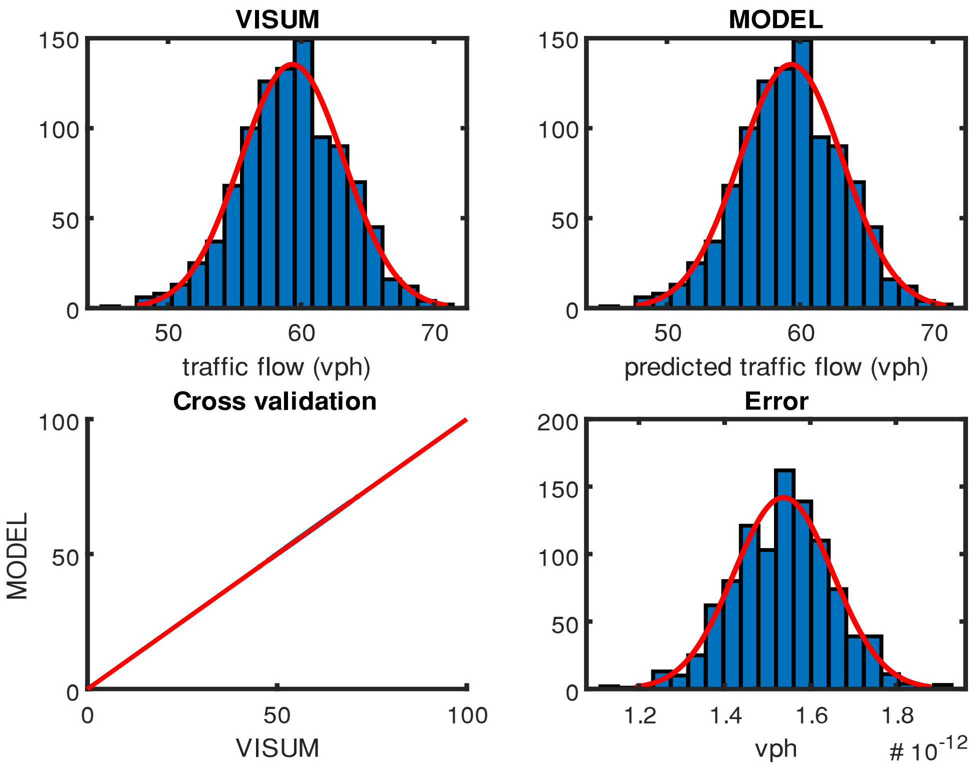


Figure 22: Show the sequential graphs for the (l_{02}) as follow: (upper-left) The calculated traffic flow by VISUM (upper-right) The predicted traffic flow by the model (lower-left) The relationship between the calculated traffic flow by VISUM and the predicted traffic flow by the model, (lower- right) The predicted error.

Table 7: Link choice set and link choice probabilities set of the link β_{12} for various RSD.

$\mathcal{F}_{02}$	β_{02}	RSD									
		0.05	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
		Link choice probabilities values									
OD_{08-01}	β_{08-01}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9986	0.9904	1.0000	1.0000
OD_{09-01}	β_{09-01}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9946	0.9956	0.9900
OD_{08-02}	β_{08-02}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988	1.0000	0.9849	1.0000	1.0000
OD_{09-02}	β_{09-02}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
OD_{06-03}	β_{06-03}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994	0.9970	1.0000	1.0000	0.9988
OD_{07-03}	β_{07-03}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9969
OD_{08-03}	β_{08-03}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000
OD_{09-03}	β_{09-03}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988	0.9939
OD_{10-03}	β_{10-03}	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	1.0000	0.9909	0.9856	1.0000
OD_{16-03}	β_{16-03}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	1.0000
OD_{17-03}	β_{17-03}	1.0000	1.0000	1.0000	1.0000	0.9997	0.9997	1.0000	1.0000	1.0000	1.0000

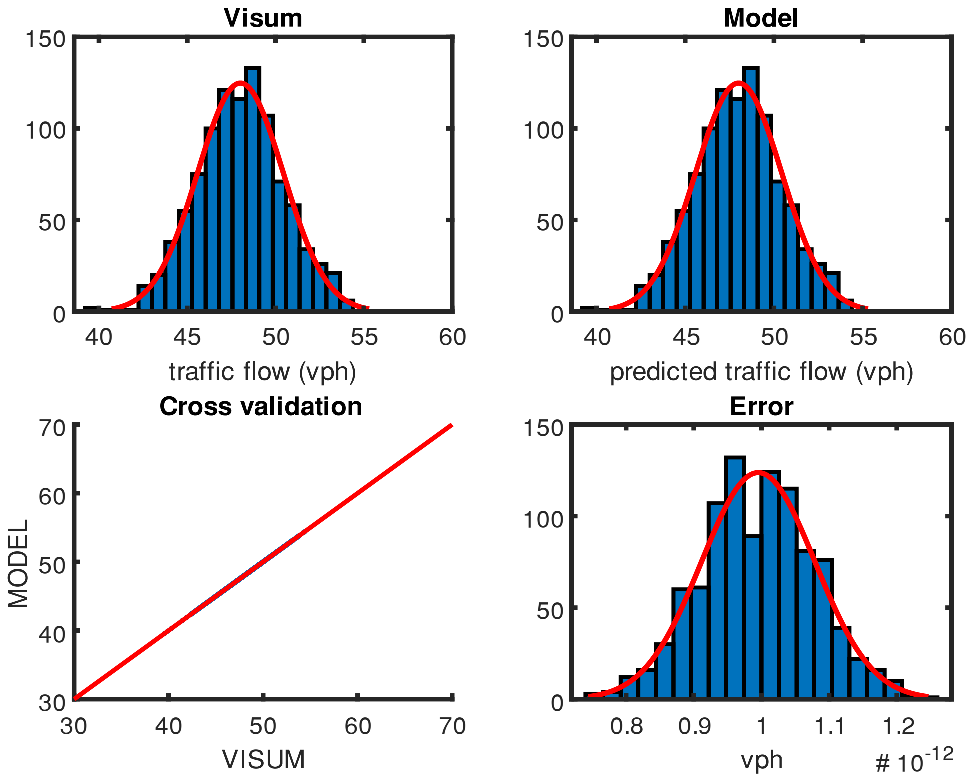


Figure 23: Show the sequential graphs for the (l_{12}) as follow: (upper-left) The calculated traffic flow by VISUM (upper-right) The predicted traffic flow by the model (lower-left) The relationship between the calculated traffic flow by VISUM and the predicted traffic flow by the model, (4) The predicted error.

It can be noticed that the results of the predicted traffic flow by the model in Equation 45 are matching the results of the calculated traffic flow by the VISUM, with extremely low errors. That proves the validity of this method to estimate choice proportion for each OD pairs in all links within the transport network.

4.5. Conclusion

This chapter presents a new methodology for route choice modelling by adopting the links as a focal and initiate point to find a choice set and estimate choice probabilities. The reason for using links instead of routes is to enable us to discover the uncertainty about the traffic flow for each link on the transport network by comparing the predicted traffic flow obtained from this methodology and the observed traffic flow.

This methodology has three key contributions:

- The Maximum likelihood ML approach was applied to determine the link choice probabilities.
- The Gaussian process GP was applied to model the prediction errors of the methodology.

The key findings provide a practical solution for link flow analysis problems. With the suggested methodology, the probability distribution function of the link flow, which makes it possible to handle model results stochastically. The method also provides a tool for estimating the error of the modelling results, which is very useful for decision-makers. The study has proved the applicability of the suggested method using a synthetic test case and a real case. The synthetic case study shows that the methodology approaches the real values very precisely with low error variances. Furthermore, the case study demonstrates the practical applicability of the proposed methodology.

Moreover, future research will be using this methodology of link choice to construct the route choice set and comparing the results with the different choice set generation methodologies.

5. Uncertainty Quantification of the Traffic Assignment Model

5.1. Introduction

Traffic assignment is a very critical part of transportation analysis. Traffic assignment models consider reconciling travel demand and transportation supply to deliver an output, which describes the steady-state of a transportation system [87]. The traffic assignment model usually possesses inherent uncertainties caused by the parametric variability of the OD matrix [16, 50]. Propagating these uncertainties through the traffic assignment model on a transport network to the outputs is one aspect of the traffic flow uncertainty [88, 89]. Beyond such propagation of OD pairs uncertainties, quantification of the output uncertainty is an essential requisite for the purpose of combine knowledge of uncertainty caused by the model structure with the uncertainty caused by the parametric variability of the OD pairs [90].

In general, the traffic assignment model's outputs are very requisite for urban and transport planners to decisions related to transport policy and infrastructure expansion. Hence, simple estimates of traffic flow, even if they refer to the most probable values, are not adequate for an adequate and safe assessment due to the risks associated with these decisions. Therefore, these planners have to give insight into the uncertainty margining the traffic assignment forecasts [24]. However, quantification of uncertainties enables the planners to know the possible range of future values for the traffic flows and the probabilities associated with these possible outputs. Accordingly, quantifying the uncertainty in traffic flows forecasts can lead to more reliable decisions [1].

Uncertainty quantification (UQ) is a technique used to quantify the uncertainty in the model output that arises from the model's parameters stochastic behaviour. The UQ depends on assigning a distribution of possible values to each model parameter instead of assuming fixed model parameters as in a deterministic model, as illustrated in Figure 24 [91]. UQ utilised both quantitative model's characterisation and the reduction of uncertainties in the models' results. UQ widely used in stretching engineering applications and applied mathematics [92].

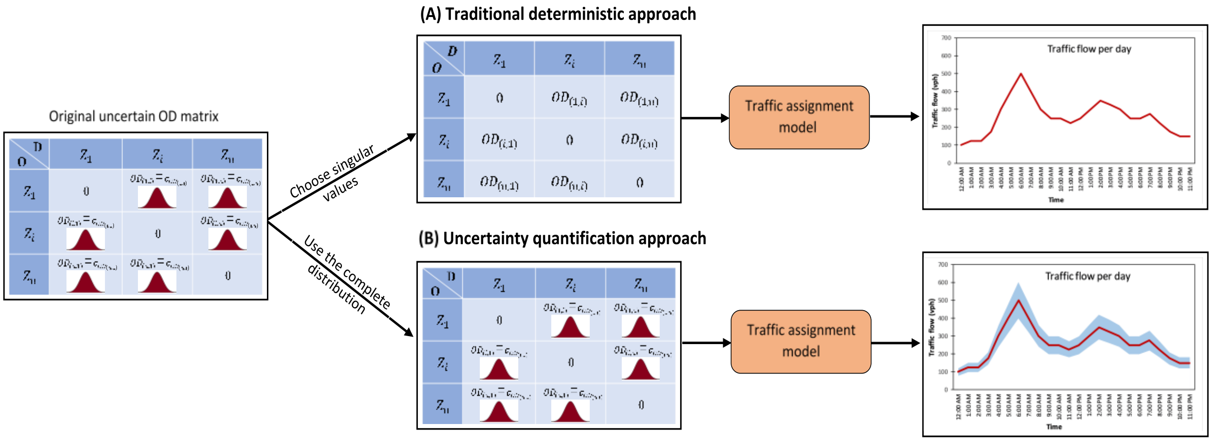


Figure 24: Illustration of uncertainty quantification of the traffic assignment model. (A) A traditional deterministic approach where each input variables have a chosen fixed value and the output of the model will be an attribute of single values. (B) uncertainty quantification approach of the traffic assignment model takes the distributions of the input variables into account, and the output of the model becomes a range of possible values.

UQ is a multidisciplinary technique with a broad base of applicable methods, including uncertainty propagation, statistical calibration, sensitivity analysis, and inverse analysis; thus, the UQ entered as a crucial ingredient in modelling workflow that aims to provide more practical information for the forecasting [1, 93]. A UQ enables us to model the naturally occurring variation in the input data (i.e., OD matrix) of the traffic assignment model. It also enhances the knowledge about the model by quantifying how the uncertain OD matrix influences traffic flow. Additionally, performing a UQ enables the transport planners to compare the outputs' variations of different traffic assignment models' methods. UQ's challenge comes from the choice of uncertain input distributions: hence, it is necessary to advance the knowledge on prior distributions and justify the choice of them. This approach is usually referred to as inverse uncertainty quantification (Inverse-UQ).

Inverse-UQ aims to estimate the statistical parameters of uncertain inputs based on given observed data [94]. It is similar to parameter calibration in that both of them use the Bayesian inference theory to explore the posterior distributions using numerical techniques [95]. However, unlike parameter calibration, Inverse-UQ also captures the uncertainty in its estimates rather than determining point estimates of the best-fit input parameters [96]. In this chapter, inverse methods based on Bayes' theorem has been used to characterise and estimate the statistical parameters of the input OD matrix of the traffic assignment model. Figure 25 shows the difference between the UQ or Forward-UQ and Inverse-UQ.

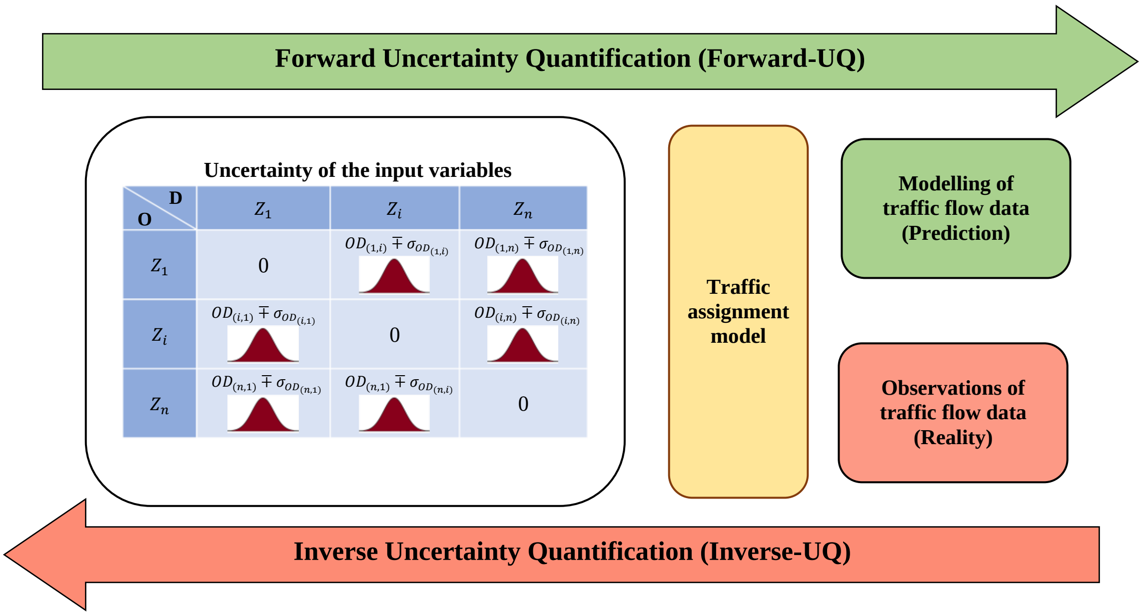


Figure 25: Illustration of the difference between the Forward-UQ and the Inverse-UQ. The Forward-UQ approach of the traffic assignment model propagates the input uncertainties through the model to predicate the traffic flow data's statistical parameters. The Inverse-UQ approach of the traffic assignment model characterises the input uncertainties to predicate the input variables' statistical parameters.

Inverse-UQ technique has gained growing attention in modelling and simulation of the scientific problems, especially in the context of source estimation and calibration of model input parameters or variables [97]. The main object of Inverse: firstly, is to address the lack of input uncertainty information issues and seek statistical descriptions of the input variables. Secondly, proper identification of the input variables' statistical parameters also yields an accurate estimation of model discrepancy outputs, which assists in discovering model deficiencies and provides guidance for future improvement of the modelling process. Thirdly, it provides a better knowledge of the calibrated parameters to produce more reliable model predictions over a comprehensive application [98, 99]. Given the fact that the exact solution of Inverse is very complex does not exist; therefore, numerical solutions usually employed to estimate the statistical space of the input parameters or variables [94, 100]. Thus, for this study, the Least Square adjustment method has been used to calibrate the OD matrix variables of the traffic assignment model.

Although many scientific articles handled the subject of uncertainty in transport models, the literature on quantifying uncertainty in the form of variances, standard deviations, confidence intervals, or other related measures is relatively limited [1]. Furthermore, the researchers who considered the quantification traffic forecasts focused on parameters of the model structure or inputs travel time and trip cost such as; Ashley [101], Lowe et al. [102], Kroes [103], Leurent [104] De Jong et al. [105], Zhao and Kockelman [2], Cools et al. [24], Deng et al. [90] and others. While the problem of the posterior distribution of the input data of the traffic assignment model did not subject to extensive studies to know the parametric variability of OD pairs except Perrakis et al. [24, 106]; they used the Markov Chain Monte Carlo method and Bayesian framework to estimate the distribution of the set of 24 regression parameters that used to generate the OD matrix of Dutch transport model. The importance of this study lies in there is an obvious need to know the leading causes that effect on fluctuation in traffic flow, which the inputs data represented by the OD matrix is one of the most significant sources of the variation Wherefore, knowing the parametric variability of input OD pairs enable us to quantify the variability of traffic flow forecasting.

In this study, a new methodology has been developed to explore posterior distributions of the input data (i.e., OD matrix) of a traffic assignment model by applying the Inverse-UQ technique. The key contributions of this methodology present:

1) Calibrating the mean values of the OD pairs to reproduce the average-OD matrix.
2) Estimating the variance values of the OD pairs to create the variance-OD matrix.
3) Quantifying the uncertainty of the traffic assignment model using both the average-OD matrix and variance-OD matrix.

5.2. Background

5.2.1. Foreword Uncertainty Quantification

Explicit Quantification of data and structural uncertainties in conceptual modelling is a major scientific and engineering challenge. In any modelling endeavour, reducing the total predictive uncertainty requires a robust quantitative understanding of each of its sources [107]. Assessment of uncertainties covers all possible outcomes and their likelihoods, not just the "most likely" outcome. Probabilistic descriptions of model inputs make it possible to derive probability distributions of model outputs and system performance indices. A quantitative approach to uncertainty analysis, where probability distribution functions are assigned to the uncertain variables, is the most appropriate way to deal with uncertainty [108].

Foreword uncertainty quantification (Foreword-UQ) is the process of representing uncertainties in system inputs and parameters, then propagating the uncertainty through the model to predict the overall uncertainty in the model results. Foreword-UQ tries to determine how likely specific outputs are if some aspects of the system are not precisely known. Moreover, Foreword-UQ describes the stochastic behaviour of outputs of interest due to uncertain inputs [109–111].

Procedures of Foreword-UQ can generally be categorised as intrusive or non-intrusive. Intrusive Foreword-UQ method requires reformulating the mathematical equations of the models describing the real system processes. Non-intrusive Foreword-UQ method, on the other hand, use sets of simulation techniques where simulation elements ensemble is generated by sampling the uncertain inputs according to different sampling schemes. The input uncertainties' impact can then be analysed for the chosen model output quantities of interest [112, 113].

This study will focus on the intrusive Foreword-UQ method; this method requires knowing the parametric variability of the input variables that describe uncertainty statistics. The errors propagation technique widely used for the Foreword-UQ problems, the basic formula for errors propagation technique as follows [114, 115]:

Let consider a model with a mathematical function:

$$y = f(x_1, x_2, \dots, x_p) \tag{47}$$

Where:(y) is the model output, $\left(x_p\right)$ being a vector of (p) input stochastic variables.

Each variable has an uncertainty represents by variance (σ_x^2). So, the mathematical function with uncertainties will be as follows:

$$y + \sigma_y^2 = f\left(\left(x_1 + \sigma_{x_1}^2\right), \left(x_2 + \sigma_{x_2}^2\right), \dots, \left(x_n + \sigma_{x_p}^2\right)\right) \tag{48}$$

Where $\left(\sigma_y^2\right)$ is the model uncertainty which, represent the variance in the model output.

The output variance $\left(\sigma_y^2\right)$, concerning the variances in variables $(x_1, x_2, \dots, x_n)$ can be approximated using partial derivatives.

$$\sigma_y^2 \simeq \sigma_{x_1}^2 \left(\frac{\partial y}{\partial x_1}\right)^2 + \sigma_{x_2}^2 \left(\frac{\partial y}{\partial x_2}\right)^2 + \cdots + \sigma_{x_p}^2 \left(\frac{\partial y}{\partial x_p}\right)^2 \tag{49}$$

This function can be applied under all conditions and can be used straight once each partial derivative is found and mathematically evaluated [115, 116].

5.2.2. Inverse Uncertainty Quantification

Inverse analysis can be used to solve uncertainty quantification problems in the mathematical and engineering sciences because of the ability to solve challenges of the interactions between analysis, computation, probability, and statistics and the applicability in various complex mathematical models. In general, analysis of the inverse problems confronts mathematical models with observed data to derive the inputs needed to run the models; information of these inputs can then be used to develop control strategies based on forecasts. Both model and input data are usually suffering to uncertainty, as are the resulting conclusions and forecasts; thus, any decisions or control strategies based on predictions will be significantly improved if quantitative uncertainty is confirmed. In this context, the I-UQ is the process of quantifying the uncertainties of input parameters based on observed outputs data [116, 117, 118].

Inverse-UQ is widely used during the last decades for complex mathematical computerised models to calibrate the uncertainty input parameters or variables by using statistical inference between the computations' results and the observed data. The calibration process is categorised into two types of deterministic calibration, and statistical calibration also called the Bayesian approach. The deterministic calibration determines the point estimates of best-fit input parameters such that the discrepancies between model output and observed data can be minimised [100, 119]. While the statistical calibration or used in Inverse-UQ concern to calibrate the parameters or variables under uncertainty produces statistical descriptions like distributions. Inverse-UQ adopts the Bayesian inference theory to explore the posterior PDF for the parameters or variables. The advantage of calibration using IUQ is apparent: firstly, information from observations is never adequately accurate to enable inference of the precise values of the input parameters or variables. Secondly, quantifying the correlations between the estimates. Finally, the observed data usually holds some degree of uncertainty, which should be considered during the inference process of calibration parameters or variables [97, 117].

The Bayesian inference theory used to determine the posterior PDF of the posterior PDF for the parameters or variables of any system model concern in the following compact form:

$$Y^{obs.}(X) = Y^{mod.}(X) + \Upsilon \tag{50}$$

Where $Y^{obs.}$ is a vector for the observations, $Y^{mod.}$ Is a vector for the system model, X is a vector for the unknown model parameters or variables, and Υ is a vector for the model uncertainty. The precise calibration parameters or variables X which is defined as $p(X|Y^{obs.}, Y^{mod.})$. According to Bayesian inference theory [120]:

$$\mathrm{P}(X|Y^{obs.}, Y^{mod.}) \propto \mathrm{P}(Y^{obs.}, Y^{mod.}|X).\mathrm{P}(X) \tag{51}$$

Where $\mathrm{P}(X)$ is the prior PDF of the parameter or variables, and $\mathrm{P}(Y^{obs.}, Y^{mod.}|X)$ is the likelihood function.

The likelihood equation that quantifies the mismatch between the model outputs $Y^{mod.}$ and the observations $Y^{obs.}$ is applied to find the posterior distribution over X:

$$L(X|Y^{obs.}, Y^{mod.}) \equiv \mathrm{P}(Y^{obs.}, Y^{mod.}|X) \tag{52}$$

When the model uncertainty is assumed to fit Gaussian distribution, the likelihood $L(X|Y^{obs.}, Y^{mod.})$ can be expressed as [83]:

$$L(X|Y^{obs.}, Y^{mod.}) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\ exp\left\{-\frac{1}{2\sigma^2}[Y^{obs.} - Y^{mod.}]^2\right\} \tag{53}$$

Where σ^2 is the errors' variance. Error ε as independent and identically distributed (i.i.d.) zero-mean Gaussian noise, whose variance is expected to be reported along with observed data. In most situations, analytical forms of $\mathrm{P}(X|Y^{obs.}, Y^{mod.})$ do not exist; therefore numerical solutions usually used to explore the input parameters or variables space and estimate $\mathrm{P}(X|Y^{obs.}, Y^{mod.})$.

5.2.3. Least Squares Adjustment

Least Squares adjustment (LSA) is a numerical technique employed to solve an overdetermined system of equations based on the principle of the Least Squares (LS) technique to minimises the sum of the squares of differences between the observation and estimate [120, 121]. The LS method is the efficient and soundest analytical technique for estimating parameters from a set of data [123]. The parameters determined by this method are normally distributed about the true parameter's value with the least possible standard deviations. The accuracy of an estimated parameter expresses the degree of closeness to its real value. Plus, the precision of the estimated parameter expresses the degree of closeness between the estimated parameter. The precision of adjusted parameters is usually given by the variance or standard deviation [124].

The principle of the LS is the estimation of unknown parameters, given a set of observations, satisfying minimises the sum of the squares of the residuals [123, 124], this can be expressed mathematically as:

$$\sum_{i=1}^{N} \Upsilon_i^2 \text{ is minimum} \tag{54}$$

Where Υ_i is the set of residuals $\Upsilon_i = \sum_{j=1}^{p} \upsilon_j^2$, p is the number of parameters.

υ_i is expressed as:

$$\upsilon_i = \mathrm{y}_i^{obs.} - \mathrm{y}_i^{mod.} \tag{55}$$

The observed values of the model system:

$$\mathrm{y}_i^{obs.} = f(X) \tag{56}$$

Where $X = x_1,\ x_2, \dots, x_p$ the set of estimated parameters, m is the number of parameters in the model.

And the approximate values of the model system:

$$\mathrm{y}_i^{mod.} = f(X^{\circ}) \tag{57}$$

Where $X^{\circ} = \mathring{x}_1,\ \mathring{x}_2, \dots, \mathring{x}_p$ is the set of approximate values of the unknown parameters.

The set of estimated parameters:

$$X = X^{\circ} + W.X^{c} \tag{58}$$

Where $X^c = x_1^c,\ x_2^c, \dots, x_p^c$ is the set of the correction to be determined by LSA, and $W = w_1, w_2, \dots, w_p$ is the weight of the parameters.

The relation between the system model and its residual to the estimated parameters and its residuals is given by:

$$\mathrm{y}_i^{mod.} + \upsilon_i = f(X^{\circ} + WX^{c}) \tag{59}$$

The function $f(X^{\circ} + WX^{c})$ can be fragmented using Taylor's series and truncating at the first order to:

$$y_i^{mod.} + v_i = f(X^{\circ}) + W\frac{\partial f}{\partial X}X^{c} \tag{60}$$

Thus, the relation between the residual system model and the set of the correction is given by:

$$v_i = W\frac{\partial f}{\partial X}X^{c} \tag{61}$$

5.2.4. Distribution Fitting

The other point in the analysis stage is fitting the distribution for the outcomes (Goodness-of-Fit Test). The distribution fitting is a procedure of choosing a statistical probability distribution that best fits the data set generated by the simulation process. Probability distributions can be used as a tool for finding the uncertainty of the model's outcomes. These distributions are offered particular indications and calculations that can apply to make well-grounded understanding about the outcomes behaviour [125].

The Goodness-Of-Fit (GOF) test describes how well the outcome set fits a set of probability distributions. The concept of GOF is comparing between two data sets: the first set is the observed values, and the second set is expected frequency, which is calculated based upon a standard distribution [126]. There are different ways to test GOF for data. In this methodology, the Kolmogorov–Smirnov test (KS test) is used to determine the type of distribution for the simulation outcomes (i.e., traffic flows).

Kolmogorov-Smirnov test (KS test):

KS test uses for the comparison of the cumulative frequencies. KS test may act as the measure of the agreement between observation and the model and a measure between two sets of observations [127].

If $F_t(X_i)$ and $F_o(X_i)$ are the theoretical and the observed cumulative frequencies for distinct and ascending ordered observations X_i, then a series of statistics are based on $F_t(X_i) - F_o(X_i)$ difference:

$$D_{-} = \max_{1<i<n}(F_t(X_i) - F_o(X_i)) \tag{62}$$

$$D_{+} = \max_{1<i<n}(F_o(X_i) - F_t(X_i)) \tag{63}$$

$$D = max(D_- , D_+) \tag{64}$$

If $D \leq KS$ critical value obtained from a table, then accept the hypothesis that data set follows the assigned distribution.

If $D > KS$ critical value obtained from a table, then reject the hypothesis that the data set follows the assigned distribution.

Where:

D_-	measures the largest leak in observation compared with theoretical value,
D_+	measures the largest excess in observation compared with theoretical value,
D	measures the largest difference in between (it is the most frequently used),
n	number of samples,

5.3. Methodology

This chapter has developed a new methodology to quantify the uncertainty in the traffic assignment model. The proposed methodology consists of two sequential processes: Inverse uncertainty quantification process and forward uncertainty quantification process. The Inverse-UQ process aims to discover the statistical parameters of the posterior distribution of the OD pairs using the Least Squares Adjustment method (LSA). While the Forward-UQ process uses the posterior distribution of the OD pairs to quantify the traffic flow forecasts. Figure 26 illustrates the methodology processes.

5.3.1. Inverse-UQ Process

For this process, the Inverse-UQ technique has been used to the statistical parameters of the posterior distribution of input OD pairs of traffic assignment model. Inverse-UQ was formulated depending on the LSA method. The Inverse-UQ Process consists of three stages:

(i) Input stage;

(ii) Calibration stage;

(iii) Calculation stage.

The relationships connecting the methodology's stages are presented in Figure 27, and the mathematical and logical computations of this methodology are illustrated in Algorithm 4.

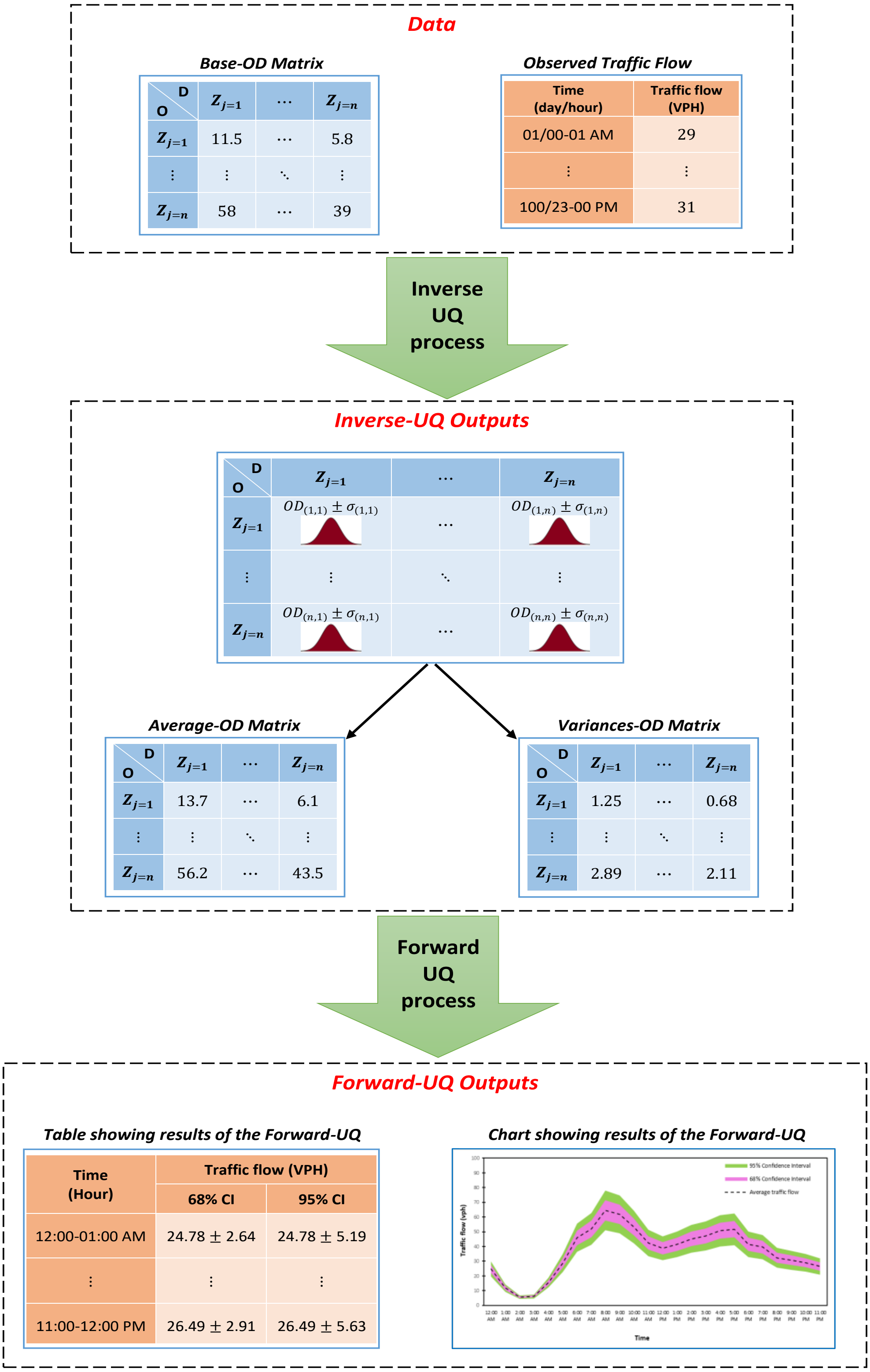


Figure 26: A flowchart illustrates the methodology process.

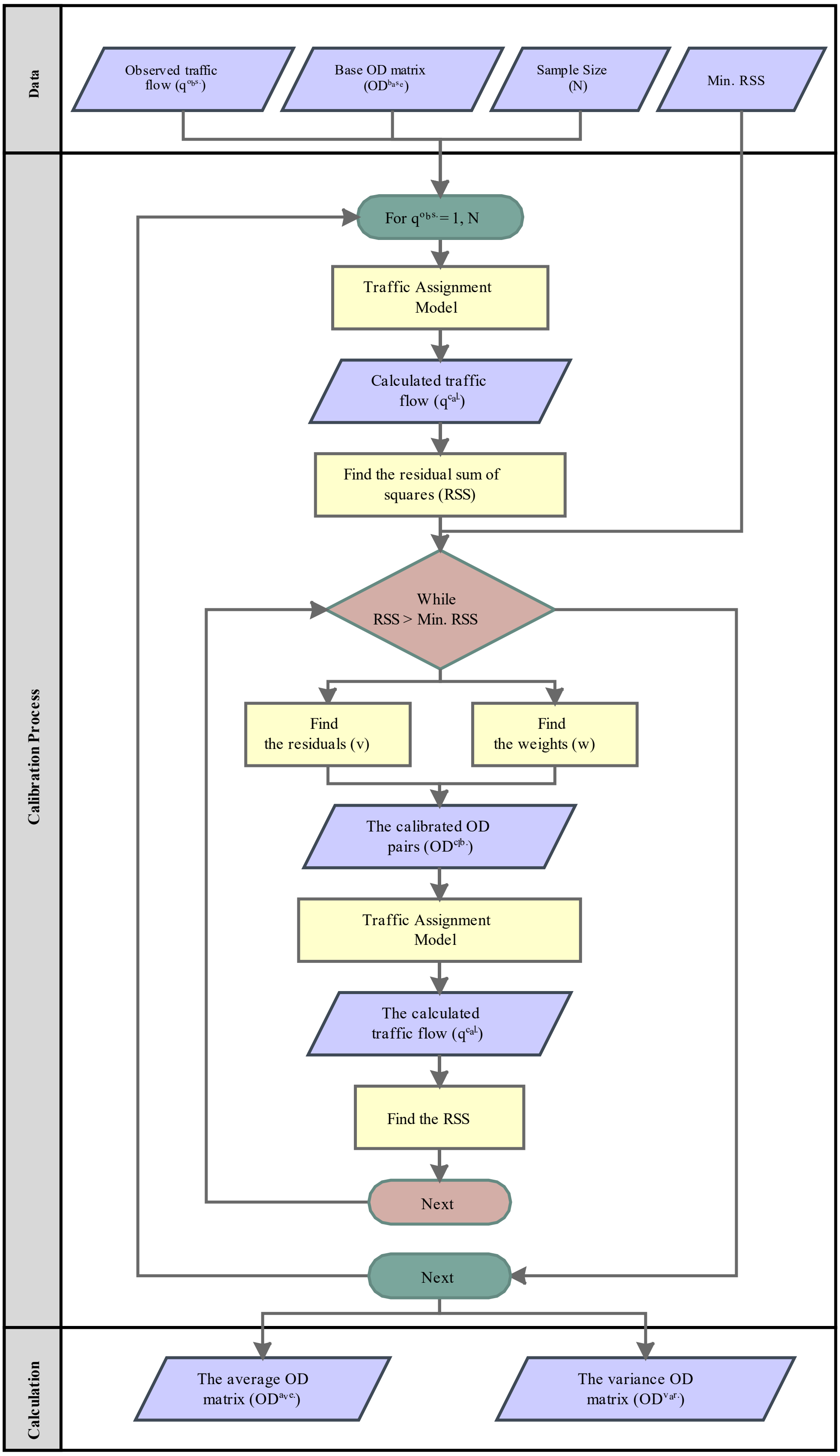


Figure 27: A flowchart illustrates the methodology stages and the relationship between the stage's components.

Algorithm 4. Inverse-UQ process.

Data stage:

$Z_{j=1,n}$; $\forall\, Z, j \in \mathbb{N}$	⊳ Zone definition.
$O_{j=1,n}$; $O \subset Z$	⊳ Origin definition.
$D_{j=1,n}$; $D \subset Z$	⊳ Destination definition.
$p = Z^2$; $p \in \mathbb{N}$	⊳ Number of OD pairs.
$OD^{obs.}_{i=1,P}$; $\forall OD^{obs.}_{i=1,P} \in \mathbb{R}$	⊳ Observed values of OD matrix.
$q^{obs.}_{l=1,L}$; $\forall q^{obs.}_{l=1,L} \in \mathbb{R}$	⊳ Observed traffic flow.
$L; L \in \mathbb{N}$	⊳ Number of links.
$\mathcal{F}_{l=1,L}$;	⊳ Link choice set.
$\beta_{l=1,L}$;	⊳ Link choice probabilites set.
$f_{T_{O\rightarrow D}}$	⊳ Trafficassignment model.
$N; N \in \mathbb{N}$	⊳ Samples size.
$Min.RSS$	⊳ Minimum residual sum of squares.

Calibration process stage:

For $k = 1\ to\ N$ **do**
- $OD^{k}_{i=1,p} = OD^{obs.}_{i=1,p}$
- $q^{cal.}_{l=1,L} \leftarrow f_{T_{O\rightarrow D}}\left\{OD^{k}_{i=1,p}\right\}$; $\forall\, q^{cal.}_{l} \in \mathbb{R}$ ⊳ Run VISUM.
- $RSS \leftarrow \sum_{l=1}^{L}(q^{obs.}_{l} - q^{cal.}_{l})^2$ ⊳ Find the residual sum of squares.
- **While** $RSS > Min.RSS$
  - **For** $l = 1\ to\ L$ **do**
    - $v_l = q^{obs.}_{l} - q^{cal.}_{l}$ ⊳ Find the residual.
    - **For** $i = 1\ to\ p$ **do**
      - $w^{l}_{i} = \left(\mathcal{F}_l * \beta_l * res * OD^{k}_{i}\right)/q^{obs.}_{l}$ ⊳ Find the weight of each OD pair.
      - $\lambda^{l}_{i} = v_l / w^{l}_{i}$ ⊳ Find the correction factor of each OD pair
    - **end for**
  - **end for**
  - **For** $i = 1\ to\ p$ **do**
    - $\lambda_i \leftarrow \left(\sum_{l=1}^{L} \lambda^{l}_{i}\right)/L$ ⊳ Find the correction factor of the OD pair
    - $OD^{k}_{i} \leftarrow OD^{k}_{i} + \lambda_i$ ⊳ Find the calibrated OD pair
  - **end for**
  - $q^{cal.}_{l=1,L} \leftarrow f_{T_{O\rightarrow D}}\left\{OD^{k}_{i=1,p}\right\}$; $\forall\, q^{cal.}_{l} \in \mathbb{R}$ ⊳ Run VISUM.
  - $RSS \leftarrow \sum_{l=1}^{L}(q^{obs.}_{l} - q^{cal.}_{l})^2$ ⊳ Find the residual sum of squares.
- **next**

end for

Results stage:

$OD^{ave.}_{i=1,p} \leftarrow \left(\sum_{k=1}^{N} OD^{k}_{i=1,p}\right)/N$ ⊳ Find the average OD matrix.

$OD^{var.}_{i=1,p} \leftarrow \left(\sum_{i=1}^{N}\left(OD^{k}_{i=1,p} - OD^{ave.}_{i=1,p}\right)^2\right)/N$ ⊳ Find the variance OD matrix.

End

5.3.1.1. Data stage

The required data for this stage consists of four main components:

- Firstly, collecting traffic flow data ($q^{obs.}$).
- Secondly, defining the base travel demand matrix (OD^{base}).
- Thirdly, specifying the sample size of observations hours (N).
- Lastly, establishing a benchmark to define the accuracy limit of the results, the methodology has adopted the Minimum residual sum of squares ($Min.RSS$). Where the RSS is a measure of the difference between the data and an estimation model. A small RSS indicates a close fit of the calibrated OD pairs to the real data.

5.3.1.2. Calibration stage

In this stage, a MATLAB code has been written and linked with the Component Object Model (COM) of VISUM software. This code was executed according to the number of samples (N). Given observed traffic flow ($q^{obs.}$), the Inverse-UQ technique based on LSA has been used to calibrate the input base OD matrix (OD^{base}). The critical components of the calibration processes are as the following steps:

1) Assuming the calibrated OD matrix ($OD^{clb.}$) equal to base OD matrix (OD^{base}) for each observed sample.

   $OD^{clb.}_{i=1,p} = OD^{base}_{i=1,p}$

2) Finding the calculated traffic flow ($q^{cal.}$), by using the input base OD matrix (OD^{base}) through the traffic assignment model in VISUM software.
3) Finding the residual sum of squares (RSS).

$$RSS = \sum_{l=1}^{L}(q_l^{obs.} - q_l^{cal.})^2 \tag{65}$$

4) Comparing the RSS with $Min.RSS$. If RSS less than or equal $Min.RSS$ that means the input OD matrix is accurate; therefore, the calibration process is not required. Otherwise, the input OD matrix needs to calibrate.
5) Finding the residual for each link on the transport network.

$$v_l = q_l^{obs.} - q_l^{cal.} \tag{66}$$

Where v_l is the residual of the traffic flow for a link (l).

6) Finding the weight of each OD pair in every link (w_i^l).

$$w_i^l = \frac{q_l^{obs.}}{\beta_l * OD_i^{clb.}} \tag{67}$$

Where β_l is the ink choice probabilities set (i.e., the proportion of the travel demand of the link choice set $\mathcal{F}_l$).

7) Find the correction factor of each OD pair in every link (λ_i^l), using the LSA method by the partial derivative for each input OD pairs in a set of link choice of the traffic assignment model.

$$\lambda_i^l = \frac{v_l}{w_i^l} \tag{68}$$

8) Finding the correction factor of each OD pair (λ_i).

$$\lambda_i = \frac{\left(\sum_{l=1}^{L} \lambda_i^l\right)}{L} \tag{69}$$

9) Finding the calibrated OD pair $\left(OD_i^{clb.}\right)$.

$$OD_i^{clb.} = OD_i^{clb.} + \lambda_i \tag{70}$$

10) Repeating steps 2 to 8.

The calibration process will continue until all the observed samples are performing.

5.3.1.3. Calculation stage

The results of the calibration stages are used to establish the average-OD matrix ($OD^{ave.}$) and the variance-OD matrix ($OD^{var.}$).

The average-OD matrix:

$$OD^{ave.}\text{matrix} = \sum_{i=1}^{p} \frac{\left(\sum_{k=1}^{N} OD_k^{clb.}\right)}{N} \tag{71}$$

The variance-OD matrix:

$$OD^{var.}\text{matrix} = \sum_{i=1}^{p} \frac{\left(\sum_{k=1}^{N}\left(OD_k^{clb.} - OD_k^{ave.}\right)^2\right)}{N} \tag{72}$$

5.3.2. Forward-UQ Process

The obtained average-OD matrix ($OD^{ave.}$) and variance-OD matrix ($OD^{var.}$) from the Inverse-UQ process will be used in the Forward-UQ process to quantify the future traffic flow forecasts. The Forward-UQ process consists of two steps; reconfiguration of the OD matrix and Quantification of the traffic flow forecasts. The mathematical computations of this methodology are illustrated in Algorithm 5.

Algorithm 5. Forward-UQ process.

Data:

$Z_{j=1,n}$; $\forall\, Z, j \in \mathbb{N}$ ⊳ Zone definition.
$O_{j=1,n}$; $O \subset Z$ ⊳ Origin definition.
$D_{j=1,n}$; $D \subset Z$ ⊳ Destination definition.
$p = Z^2$; $P \in \mathbb{N}$ ⊳ Number of OD pairs.
$f_{T_{O \to D}}$ ⊳ Trafficassignment model.
$OD_{(n,n)}^{ave.}$; $\forall OD_{(n,n)}^{ave.} \in \mathbb{R}$ ⊳ Average-OD matrix.
$OD_{(n,n)}^{var.}$; $\forall OD_{(n,n)}^{var.} \in \mathbb{R}$ ⊳ Variances-OD matrix.

Reconfiguration:

01: **For** $j = 1\ to\ n$ **do**
02: **For** $j = 1\ to\ n$ **do**
03: $OD_{(j,j)}^{max.prob.} @CI = 68\% = OD_{(j,j)}^{ave.} + 1.00\sqrt{OD_{(j,j)}^{var.}}$ ⊳ Find the max. probable OD matrix @ CI=68%.
04: $OD_{(j,j)}^{min.prob.} @CI = 68\% = OD_{(j,j)}^{ave.} - 1.00\sqrt{OD_{(j,j)}^{var.}}$ ⊳ Find the min. probable OD matrix @ CI=68%.
05: $OD_{(j,j)}^{max.prob.} @CI = 95\% = OD_{(j,j)}^{ave.} + 1.96\sqrt{OD_{(j,j)}^{var.}}$ ⊳ Find the max. probable OD matrix @ CI=95%.
06: $OD_{(j,j)}^{min.prob.} @CI = 95\% = OD_{(j,j)}^{ave.} - 1.96\sqrt{OD_{(j,j)}^{var.}}$ ⊳ Find the min. probable OD matrix @ CI=95%.
07: **end for**
08: **end for**

Quantification:

09: $q_l^{ave.} = f_{T_{O \to D}}\left(OD^{ave.}\right)$ ⊳ Run VISUM to find the average traffic flow.
10: $q_l^{max.prob.} @CI = 68\% = f_{T_{O \to D}}\left(OD_{(n,n)}^{max.prob.} @CI = 68\%\right)$ ⊳ Run VISUM to find the max. probable traffic flow @ CI=68%.
11: $q_l^{min.prob.} @CI = 68\% = f_{T_{O \to D}}\left(OD_{(n,n)}^{min.prob.} @CI = 68\%\right)$ ⊳ Run VISUM to find the min. probable traffic flow @ CI=68%.
12: $q_l^{max.prob.} @CI = 95\% = f_{T_{O \to D}}\left(OD_{(n,n)}^{max.prob.} @CI = 95\%\right)$ ⊳ Run VISUM to find the max. probable traffic flow @ CI=95%.
13: $q_l^{min.prob.} @CI = 95\% = f_{T_{O \to D}}\left(OD_{(n,n)}^{min.prob.} @CI = 95\%\right)$ ⊳ Run VISUM to find the min. probable traffic flow @ CI=95%.
14: **End**

5.3.2.1. Reconfiguration the OD matrix

both of the average-OD matrix ($OD^{ave.}$) and the variance-OD matrix ($OD^{var.}$) are utilised to configure the maximum probable-OD matrix ($OD^{max.pob.}$) and minimum probable-OD matrix ($OD^{min.pob.}$) according to different confidence intervals (CI). For this study, two confidence intervals have been adopted (CI): 68% and 95%, The $CI = 68\%$ comparable to $1.00 \times \sqrt{\sigma^2}$, on the other hand, the $CI = 95\%$ comparable to $1.96 \times \sqrt{\sigma^2}$.

The max. probable OD matrix within (68% CI):

$$OD^{max.prob.@CI=68\%} = OD^{ave.} + 1.00 \times \sqrt{OD^{var.}} \tag{73}$$

The min. probable OD matrix within (68% CI):

$$OD^{min.prob.@CI=68\%} = OD^{ave.} - 1.00 \times \sqrt{OD^{var.}} \tag{74}$$

The max. probable OD matrix within (95% CI):

$$OD^{max.prob.@CI=95\%} = OD^{ave.} + 1.96 \times \sqrt{OD^{var.}} \tag{75}$$

The min. probable OD matrix within (95% CI):

$$OD^{min.prob.@CI=95\%} = OD^{ave.} - 1.96 \times \sqrt{OD^{var.}} \tag{76}$$

Furthermore, the bias calculation for every base-OD pair:

$$\eta = OD^{base} - OD^{ave.} \tag{77}$$

5.3.2.2. Quantification the traffic flow forecasts

Finally, quantifying the traffic flow in each link on the transport network, using the errors propagation theory, by finding the average traffic flow ($q_l^{ave.}$) and the probable flow $\left(q_l^{prob.}\right)$.

The average traffic flow ($q_l^{ave.}$):

$$q_l^{ave.} = f_{T_{O\rightarrow D}}\left(OD^{ave.}\right) \tag{78}$$

The max. probable traffic flow value of at $CI = 68\%$:

$$q_l^{max.prob.@CI=68\%} = f_{T_{O\rightarrow D}}\left(OD^{max.prob.@CI=68\%}\right) \tag{79}$$

The min. probable traffic flow value of at $CI = 68\%$:

$$q_l^{min.prob.@CI=68\%} = f_{T_{O\rightarrow D}}\left(OD^{min.prob.@CI=68\%}\right) \tag{80}$$

The max. probable traffic flow value of at $CI = 95\%$:

$$q_l^{max.prob.@CI=95\%} = f_{T_{O\rightarrow D}}\left(OD^{max.prob.@CI=95\%}\right) \tag{81}$$

The min. probable traffic flow value of at $CI = 95\%$:

$$q_l^{min.prob.@CI=95\%} = f_{T_{O \to D}} \left(OD^{min.prob.@CI=95\%} \right) \tag{82}$$

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

Moreover, for the validation purpose to the accuracy of methodology results, the methodology can compare the estimated traffic flow by this methodology $(q_l^{est.})$ with the real observed traffic flow $\left(q_l^{obs.}\right)$.

The estimated traffic flow $(q_l^{est.})$:

$$q_l^{est.} = f_{T_{O \to D}} \left(OD^{clb.} \right) \tag{83}$$

5.4. Implementation

The proposed methodology has been applied to the case study (Ajka, Hungary). The following subsections represent the implementation stages:

The inputs data for this study case include base-OD matrix as shown in Table 8, the traffic flow observation period continued for 100 days for all transportation links (i.e., sample size N=100), and the $min. RSS \leq 1$. The final result of the calibration process is represented into two matrices; the first is the average-OD matrix $(OD^{ave.})$ as shown in Table 9, and the second is the variance-OD matrix $(OD^{var.})$, as shown in Table 10. These two matrices providing full information about the probability distributions for all OD pairs by giving the mean (μ) and the variance (σ^2) for each one.

Table 8: The base-OD matrix.

O \ D	Z-01	Z-02	Z-03	Z-04	Z-05	Z-06	Z-07	Z-08	Z-09	Z-10	Z-11	Z-12	Z-13	Z-14	Z-15	Z-16	Z-17	Z-18	Z-19	Z-20	Z-21	Z-22	Z-23	Z-24	Z-25
Z-01	0.0	30.1	263.2	7.2	0.6	6.7	4.1	5.6	8.2	6.1	9.2	6.1	5.6	5.6	3.6	3.1	3.6	3.6	6.1	6.1	3.6	3.1	13.8	9.7	7.2
Z-02	9.1	0.0	8.6	0.2	0.0	0.2	0.1	0.2	0.3	0.2	0.3	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.5	0.3	0.2
Z-03	268.6	29.1	0.0	6.9	0.6	6.4	4.0	5.4	7.9	5.9	8.9	5.9	5.4	5.4	3.5	3.0	3.5	3.5	5.9	5.9	3.5	3.0	13.4	9.4	6.9
Z-04	6.8	0.7	6.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-05	0.5	0.1	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-06	6.3	0.7	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-07	3.9	0.4	3.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-08	5.4	0.6	5.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-09	7.8	0.8	7.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-10	5.8	0.6	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-11	8.8	0.9	8.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-12	5.8	0.6	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-13	5.4	0.6	5.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-14	5.4	0.6	5.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-15	3.4	0.4	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-16	2.9	0.3	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-17	3.4	0.4	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-18	3.4	0.4	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-19	5.8	0.6	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-20	5.8	0.6	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-21	3.4	0.4	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-22	2.9	0.3	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-23	13.1	1.4	12.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-24	9.2	1.0	8.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-25	6.8	0.7	6.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 9: The average-OD matrix.

O \ D	Z-01	Z-02	Z-03	Z-04	Z-05	Z-06	Z-07	Z-08	Z-09	Z-10	Z-11	Z-12	Z-13	Z-14	Z-15	Z-16	Z-17	Z-18	Z-19	Z-20	Z-21	Z-22	Z-23	Z-24	Z-25
Z-01	0.0	31.6	292.6	9.3	0.8	7.6	4.4	5.6	9.3	7.2	9.8	7.4	6.2	6.9	3.9	3.5	3.3	3.3	11.9	5.8	7.2	3.4	14.5	11.5	7.2
Z-02	11.8	0.0	9.6	0.3	0.0	0.2	0.1	0.2	0.3	0.2	0.3	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.4	0.2	0.2	0.1	0.5	0.5	0.3
Z-03	301.1	32.7	0.0	8.6	0.7	7.0	4.1	6.3	7.8	6.2	9.7	6.7	6.1	5.9	3.9	3.9	3.5	3.9	7.0	6.7	0.8	3.0	15.0	10.5	7.8
Z-04	7.5	0.8	7.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-05	0.5	0.1	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-06	6.8	0.7	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-07	4.6	0.5	4.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-08	5.8	0.6	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-09	8.5	0.9	8.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-10	6.7	0.7	6.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-11	8.7	1.0	8.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-12	7.0	0.8	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-13	5.9	0.6	5.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-14	6.0	0.6	5.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-15	3.8	0.4	3.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-16	3.1	0.3	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-17	3.5	1.5	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-18	3.3	0.4	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-19	3.1	0.2	4.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-20	5.8	0.6	6.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-21	5.3	0.6	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-22	3.3	0.4	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-23	17.0	1.4	13.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-24	9.5	1.0	9.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z-25	6.8	0.8	7.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 10: The variance-OD matrix.

O \ D	Z-01	Z-02	Z-03	Z-04	Z-05	Z-06	Z-07	Z-08	Z-09	Z-10	Z-11	Z-12	Z-13	Z-14	Z-15	Z-16	Z-17	Z-18	Z-19	Z-20	Z-21	Z-22	Z-23	Z-24	Z-25
Z-01	0.00	10.86	981.5	0.95	0.01	0.76	0.22	0.37	1.05	0.76	1.18	0.64	0.48	0.57	0.19	0.19	0.14	0.33	1.91	0.44	1.02	0.13	2.29	1.48	0.00
Z-02	1.65	0.00	1.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-03	1035	12.31	0.00	0.86	0.01	0.60	0.22	0.57	0.73	0.49	1.17	0.51	0.46	0.41	0.19	0.23	0.14	0.18	0.65	0.51	0.14	0.00	2.60	1.26	0.69
Z-04	0.66	0.01	0.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-06	0.51	0.01	0.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-07	0.31	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-08	0.39	0.00	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-09	0.89	0.01	0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-10	0.64	0.01	0.54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-11	0.75	0.01	0.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-12	0.55	0.01	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-13	0.42	0.00	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-14	0.37	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-15	0.17	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-16	0.14	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-17	0.13	0.24	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-18	0.13	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-19	0.25	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-20	0.46	0.01	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-21	0.58	0.01	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-22	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-23	3.42	0.00	2.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-24	1.02	0.01	1.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z-25	0.00	0.01	0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

For this case study, the application of the proposed methodology has been presented for one link only; the chosen link is (l_{02}), this link summarises the traffic assignment pattern of other links in this case study. Figure 28 shows the traffic flow (q_{02}) at the time (8:00-9:00 AM) during observed periods. The link choice set of link #2 ($\mathcal{F}_{02}$) has involved 17 OD pairs, and the link choice probabilities set is 1.00 for all pairs.

$$\mathcal{F}_{02} = \{OD_{04-01}, OD_{05-01}, OD_{06-01}, OD_{07-01}, OD_{11-01}, OD_{12-01}, OD_{16-01}, OD_{04-02}, OD_{06-02}, OD_{07-02}, OD_{11-02}, OD_{12-02}, OD_{16-02}, OD_{04-03}, OD_{05-03}, OD_{11-03}, OD_{12-03}\}$$

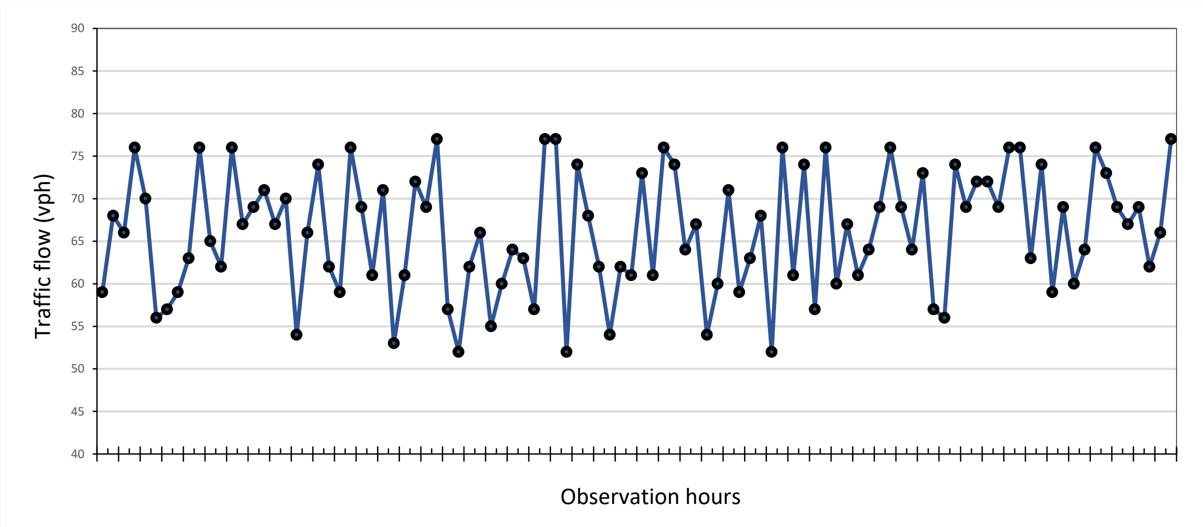


Figure 28: Traffic flow (q_{02}) at the time (8:00-9:00 AM) during the observed period.

The results of the (Goodness-of-Fit Test) for the case study (Ajka) show that:

- 98 % of OD pairs follow Normal distribution under significance level = 0.05,
- 95 % of OD pairs follow Normal distribution under significance level = 0.10,
- 91 % of OD pairs follow Normal distribution under significance level = 0.20,

For example, Table 11 and Table 12 illustrate the Kolmogorov-Smirnov test for the (OD_{13-01}) and (OD_{06-01}), respectively. Furthermore, both Figure 29 and Figure 30 are showing the (Goodness-of-Fit Test) for the (OD_{13-01}) and (OD_{06-01}), respectively.

And, The result of the calibration process for the set ($\mathcal{F}_{02}$) illustrated in Figure 31, which shows the scatter-plots of the calibrated-OD pairs ($OD^{clb.}$) vs the observed traffic flow ($q_2^{obs.}$) surrounded by 68% and 95% confidence intervals boundaries, and Figure 32, which shows probability distributions of the OD pairs of the set $\mathcal{F}_{02}$.

Table 11: Kolmogorov-Smirnov Test for the (OD_{13-01})

Kolmogorov-Smirnov Test					
Sample Size (n) Statistic (D)	100 0.073				
Significance level (α)	0.2	0.1	0.05	0.02	0.01
KS critical value	0.106	0.121	0.134	0.15	0.161
Reject?	No	No	No	No	No

Table 12: Kolmogorov-Smirnov Test for the ($OD_{06\text{-}01}$)

Kolmogorov-Smirnov Test					
Sample Size (n)	100				
Statistic (D)	0.131				
Significance level (α)	0.2	0.1	0.05	0.02	0.01
KS critical value	0.106	0.121	0.134	0.15	0.161
Reject?	Yes	Yes	No	No	No

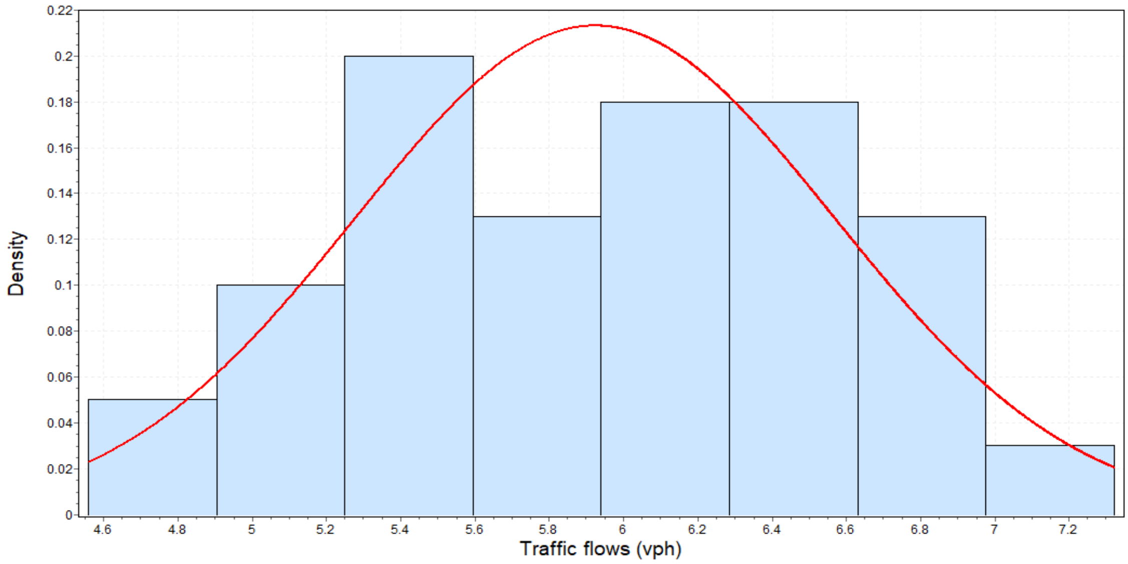


Figure 29: The Goodness-of-Fit Test for the ($OD_{13\text{-}01}$)

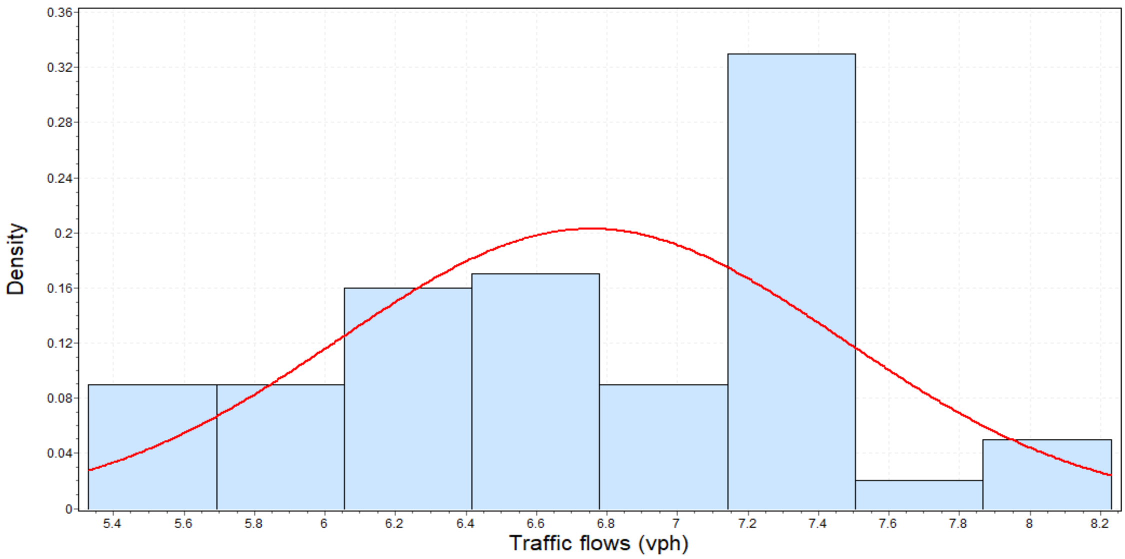


Figure 30: The Goodness-of-Fit Test for the ($OD_{06\text{-}01}$)

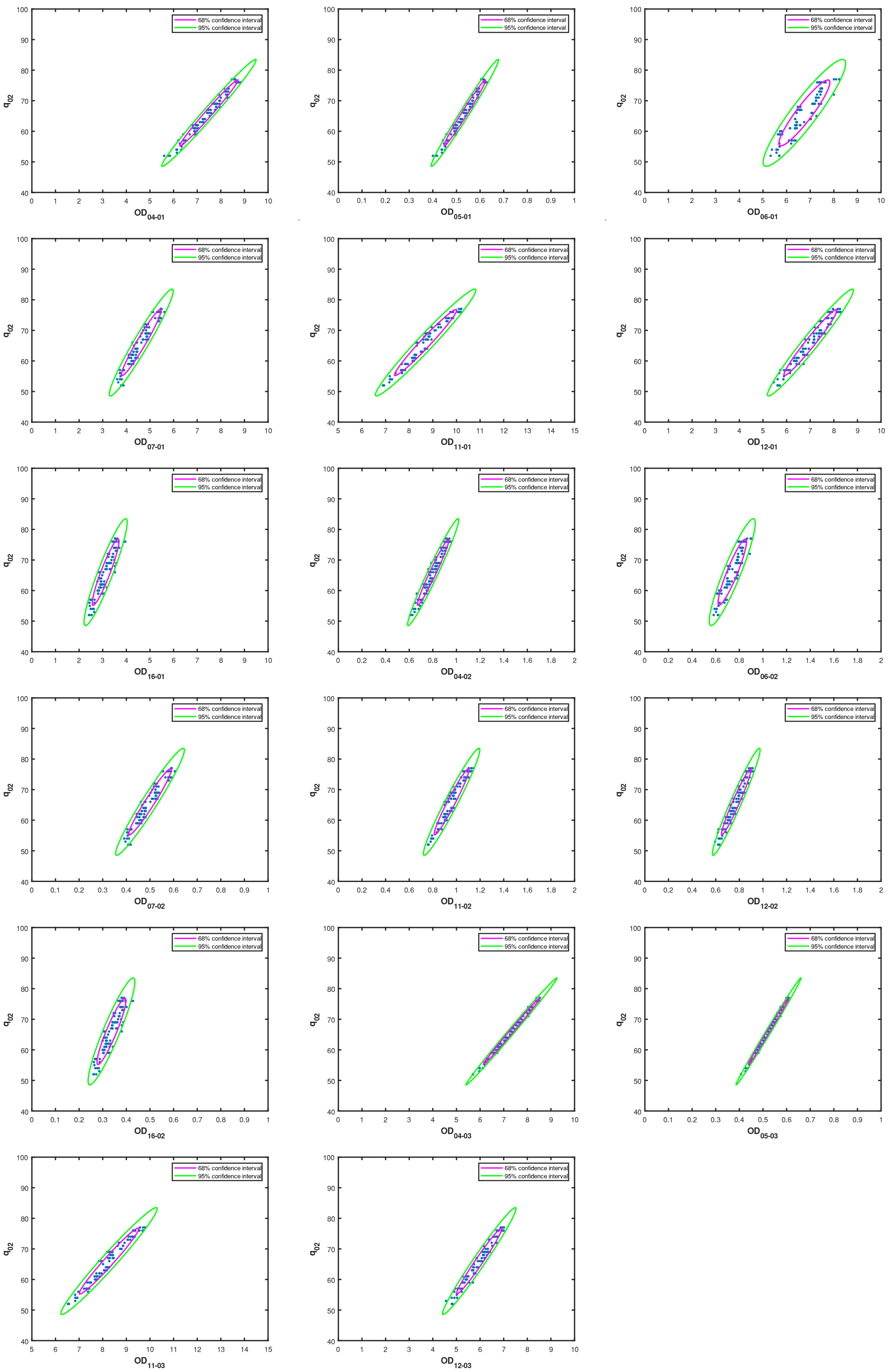


Figure 31: Scatter-plots of the calibrated-OD pairs ($OD^{clb.}$) vs the observed traffic flow ($q_2^{obs.}$), and surrounded by 68% and 95% confidence intervals boundaries.

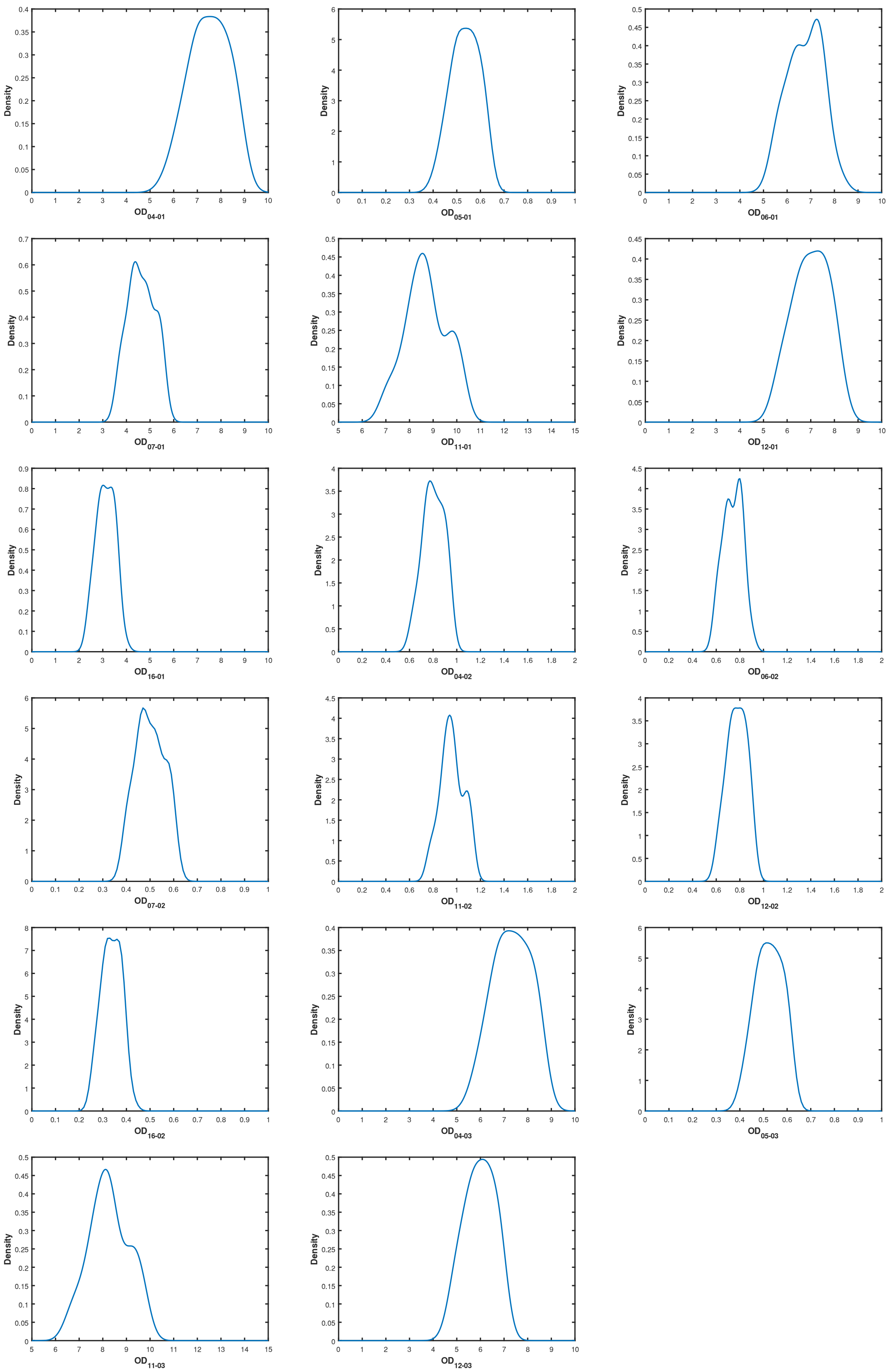


Figure 32: Probability distributions of the OD pairs of the set ($\boldsymbol{\mathcal{F}_{02}}$).

To provide reliable insight into the accuracy of the calibration process. A comparison has been created between the observed traffic flow $\left(q_2^{obs.}\right)$ with the estimated traffic flow $(q_2^{est.})$ produced by the traffic assignment model via VISUM software using the calibrated-OD matrix $(OD^{clb.})$. The comparison was repeated 100 times, according to the number of observations (N). Figure 33 presents the comparison result, the right-side showing the probability distributions of both observed traffic flow $\left(q_2^{obs.}\right)$ and estimated traffic flow $(q_2^{est.})$, It can be seen a significant matching between them. Moreover, the right-side indicating the cross-validation of observed traffic flow $\left(q_2^{obs.}\right)$ and estimated traffic flow $(q_2^{est.})$, which gives $R^2 = 0.9994$.

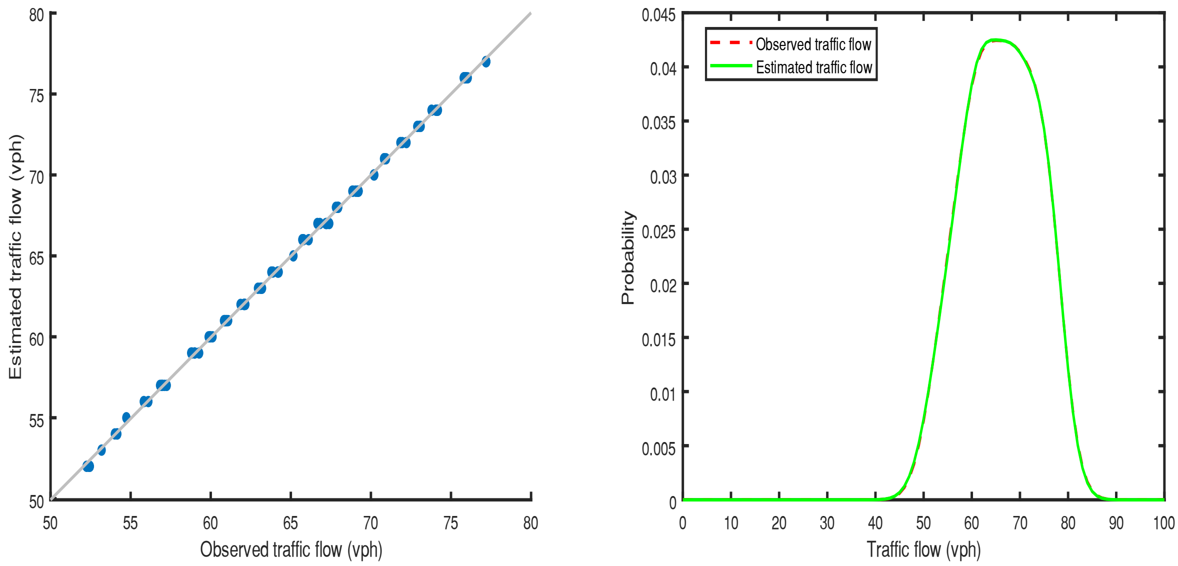


Figure 33: Illustrating the comparison between the observed traffic flow $\left(\boldsymbol{q_2^{obs.}}\right)$ and estimated traffic flow $(\boldsymbol{q_2^{est.}})$, the right-side shows probability distributions of them and the left-side shows the cross-validation between them.

For the sake of simplicity, the statistical parameters of the $\mathcal{F}_{02}$ set with their analytical conclusions are illustrated in Table 13, which includes the base-OD pairs (OD^{base}), the distribution parameters of OD pairs: average-OD pairs $(OD^{ave.})$ and variances-OD pairs $(OD^{var.})$. As well as, bias (η) calculations of the base-OD pairs (OD^{base}) compared to the average-OD pairs $(OD^{ave.})$, and the probable-OD pairs $(OD^{prob.})$, under two confidence intervals, 68% and 95%, respectively. It can be noticed in this case study that the most base-OD pairs (OD^{base}) were having a negative bias from the average-OD values $(OD^{ave.})$, except OD_{11-01} and OD_{11-03} which they had a positive bias.

Table 13: Illustrated the statistical parameters of the $\mathcal{F}_{02}$ set.

Link choice set ($\mathcal{F}_{02}$)	Base travel demand ($\boldsymbol{OD}^{base}$)	Average travel demand ($\boldsymbol{OD}^{ave.}$)	Travel demand variances ($\boldsymbol{OD}^{var.}$)	Bias (η)	Probable travel demand ($\boldsymbol{OD}^{pob.}_{CI=68\%}$)	Probable travel demand ($\boldsymbol{OD}^{pob.}_{CI=95\%}$)
$\boldsymbol{OD_{04-01}}$	6.81	7.50	0.664	-0.69	7.50 ± 0.815	7.50 ± 1.597
$\boldsymbol{OD_{05-01}}$	0.49	0.54	0.003	-0.05	0.54 ± 0.055	0.54 ± 0.107
$\boldsymbol{OD_{06-01}}$	6.33	6.75	0.509	-0.42	6.75 ± 0.713	6.75 ± 1.398
$\boldsymbol{OD_{07-01}}$	3.89	4.63	0.305	-0.74	4.63 ± 0.552	4.63 ± 1.082
$\boldsymbol{OD_{11-01}}$	8.76	8.69	0.755	+0.07	8.69 ± 0.869	8.69 ± 1.703
$\boldsymbol{OD_{12-01}}$	5.84	7.01	0.552	-1.17	7.01 ± 0.743	7.01 ± 1.456
$\boldsymbol{OD_{16-01}}$	2.92	3.12	0.139	-0.20	3.12 ± 0.373	3.12 ± 0.731
$\boldsymbol{OD_{04-02}}$	0.74	0.80	0.008	-0.06	0.80 ± 0.089	0.80 ± 0.175
$\boldsymbol{OD_{06-02}}$	0.69	0.74	0.006	-0.05	0.74 ± 0.077	0.74 ± 0.152
$\boldsymbol{OD_{07-02}}$	0.42	0.50	0.004	-0.08	0.50 ± 0.063	0.50 ± 0.124
$\boldsymbol{OD_{11-02}}$	0.95	0.96	0.010	-0.01	0.96 ± 0.100	0.96 ± 0.196
$\boldsymbol{OD_{12-02}}$	0.63	0.77	0.007	-0.14	0.77 ± 0.084	0.77 ± 0.164
$\boldsymbol{OD_{16-02}}$	0.32	0.34	0.002	-0.02	0.34 ± 0.045	0.34 ± 0.088
$\boldsymbol{OD_{04-03}}$	6.45	7.33	0.617	-0.88	7.33 ± 0.785	7.33 ± 1.540
$\boldsymbol{OD_{05-03}}$	0.46	0.52	0.003	-0.06	0.52 ± 0.055	0.52 ± 0.107
$\boldsymbol{OD_{11-03}}$	8.29	8.27	0.694	+0.02	8.27 ± 0.833	8.27 ± 1.633
$\boldsymbol{OD_{12-03}}$	5.53	5.97	0.403	-0.44	5.97 ± 0.635	5.97 ± 1.244

Finally, the traffic flow uncertainty has been quantified by including the variations of the OD pairs of the set ($\mathcal{F}_{02}$) by finding the probable traffic flow ($q_2^{prob.}$) in confidence intervals: 68% and 95%, respectively. The results of the Foreword-UQ are illustrated with details in Figure 35 and Table 14.

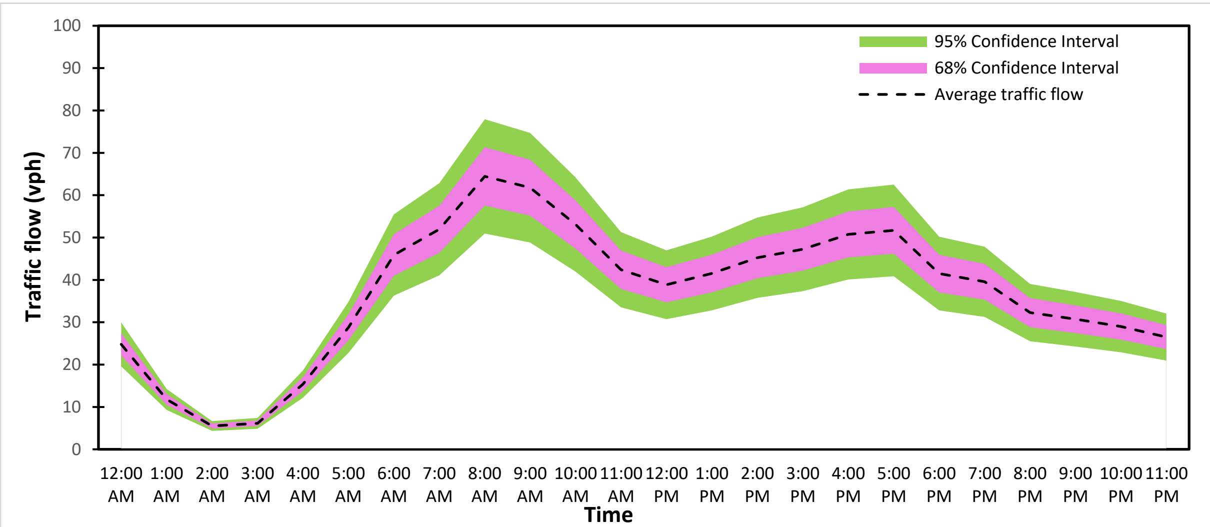


Figure 32: Showing the average traffic flow and the probable traffic flow at 68% and 95% confidence interval in link #2.

Table 14: Illustrated the average traffic flow and the probable traffic flow at 68% and 95% confidence interval in link #2.

Time	Average traffic flow $\left(q_2^{ave.}\right)$	68% CI	95% CI
		Probable traffic flow $(q_2^{prob.})$	Probable traffic flow $(q_2^{prob.})$
12:00-01:00 AM	24.78	24.78 ± 2.64	24.78 ± 5.19
01:00-02:00 AM	11.78	11.78 ± 1.36	11.78 ± 2.57
02:00-03:00 AM	5.51	5.51 ± 0.99	5.51 ± 1.05
03:00-04:00 AM	6.14	6.14 ± 0.86	6.14 ± 1.29
04:00-05:00 AM	15.40	15.40 ± 1.65	15.40 ± 2.63
05:00-06:00 AM	28.77	28.77 ± 3.37	28.77 ± 6.09
06:00-07:00 AM	45.88	45.88 ± 4.79	45.88 ± 9.71
07:00-08:00 AM	51.96	51.96 ± 5.51	51.96 ± 10.98
08:00-09:00 AM	64.44	64.44 ± 6.89	64.44 ± 13.80
09:00-10:00 AM	61.78	61.78 ± 6.60	61.78 ± 12.94
10:00-11:00 AM	53.08	53.08 ± 5.67	53.08 ± 11.02
11:00-12:00 AM	42.39	42.39 ± 4.53	42.39 ± 8.80
12:00-01:00 PM	38.85	38.85 ± 4.05	38.85 ± 8.19
01:00-02:00 PM	41.52	41.52 ± 4.54	41.52 ± 8.55
02:00-03:00 PM	45.23	45.23 ± 4.63	45.23 ± 9.67
03:00-04:00 PM	47.24	47.24 ± 5.15	47.24 ± 9.79
04:00-05:00 PM	50.74	50.74 ± 5.26	50.74 ± 10.83
05:00-06:00 PM	51.69	51.69 ± 5.32	51.69 ± 11.01
06:00-07:00 PM	41.48	41.48 ± 4.43	41.48 ± 8.69
07:00-08:00 PM	39.56	39.56 ± 4.35	39.56 ± 8.05
08:00-09:00 AM	32.28	32.28 ± 3.65	32.28 ± 6.96
09:00-10:00 PM	30.71	30.71 ± 3.21	30.71 ± 6.33
10:00-11:00 PM	28.98	28.98 ± 3.16	28.98 ± 6.17
11:00-12:00 PM	26.49	26.49 ± 2.91	26.49 ± 5.63

5.5. Conclusions

This chapter introduces a practical methodology based on the UQ framework for quantifying the traffic flow uncertainty of a traffic assignment model. The methodology aims to explore the uncertainties of the OD matrix and estimate their parameters, which cause a variation and uncertainty in the traffic flow of the traffic assignment model. The framework implemented in this methodology principally depends on the Inverse-UQ technique based on the least-squares adjustment (LSA) method used to calibrate and estimate the distribution parameters (mean and variance) of all elements of the OD matrix. Accordingly, the key contributions of the proposed methodology present:

(1) Calibration of the mean values of the OD pairs to reproduce an average-OD matrix.

(2) Estimation of the variances of the OD pairs to create a variance-OD matrix.

(3) Quantifying the uncertainty of the traffic assignment model using both average-OD and variance-OD matrices.

The methodology was then applied successfully in a case study (Ajka, Hungary). The inputs of the base-OD matrix and the observed traffic flow were used in the Inverse-UQ framework to estimate OD pairs' posterior distributions. The posterior distributions were then employed in Forward-UQ to quantify the traffic flow uncertainty for all links on the transport network. Hence, I strongly advised using this methodology because it enables the decision-makers to expect traffic flow forecasting variations.

6. New Scientific Results

In this chapter, the main new scientific results are summarised in the form of the following theses. Next, a short outlook is given about corresponding research objectives for the future. Following, the suggested visions of research applications. Finally, the list of publications related to theses.

6.1. Theses

The new scientific results are written in bold, while the short explanations are written in normal characters. The numbers of the referring publications can be seen in square brackets.

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.

This methodology was built using the Monte Carlo simulation method to generate various OD matrices for the traffic assignment model via VISUM software to produce traffic flow on a transport network. The values of the OD matrix were considered as stochastic variables following different probability distributions. The results of the simulation process present the predictive traffic flow in each link on the transport network.

Furthermore, the uncertainty analysis of the observed traffic flow was based on the bias-variance decomposition method. The bias error represents the difference between the average traffic flow prediction and the observed traffic flow. Simultaneously, the variance represents the variability in the traffic assignment model prediction due to the stochastic nature of the input data. The bias and variability of the predictive traffic flow show the uncertainty propagation through the traffic assignment model.

Subsequently, the uncertainty analysis results are classified into four cases according to bias and variance characterisations for each link on the transport network. The proposed classification was consisting of four cases of uncertainty: Case I (low variability - low bias), Case II (low variability - high bias), Case III (high variability - low bias), and Case IV (high variability - high bias). Consequently, the uncertainty cases are visualised in coloured figures to clarify the likelihood of occurring for all transport links.

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

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 GSA methods, and traffic assignment model TAM via VISUM software. 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.

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

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 to fill the gap of knowledge about the proportions of the OD pairs of the link choice set 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. This method's importance 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.

Related publication: [J-2]

Thesis IV:

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

This method's key contributions represent 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 output 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 and MC simulation method,

(3) MC simulation method,

(4) counting traffic flow data.

Related publication: [J-3]

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. The result of this methodology enables us to quantify the uncertainty of predictive traffic flow.

Related publication: [J-3]

6.2. 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.

6.3. 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 33 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.

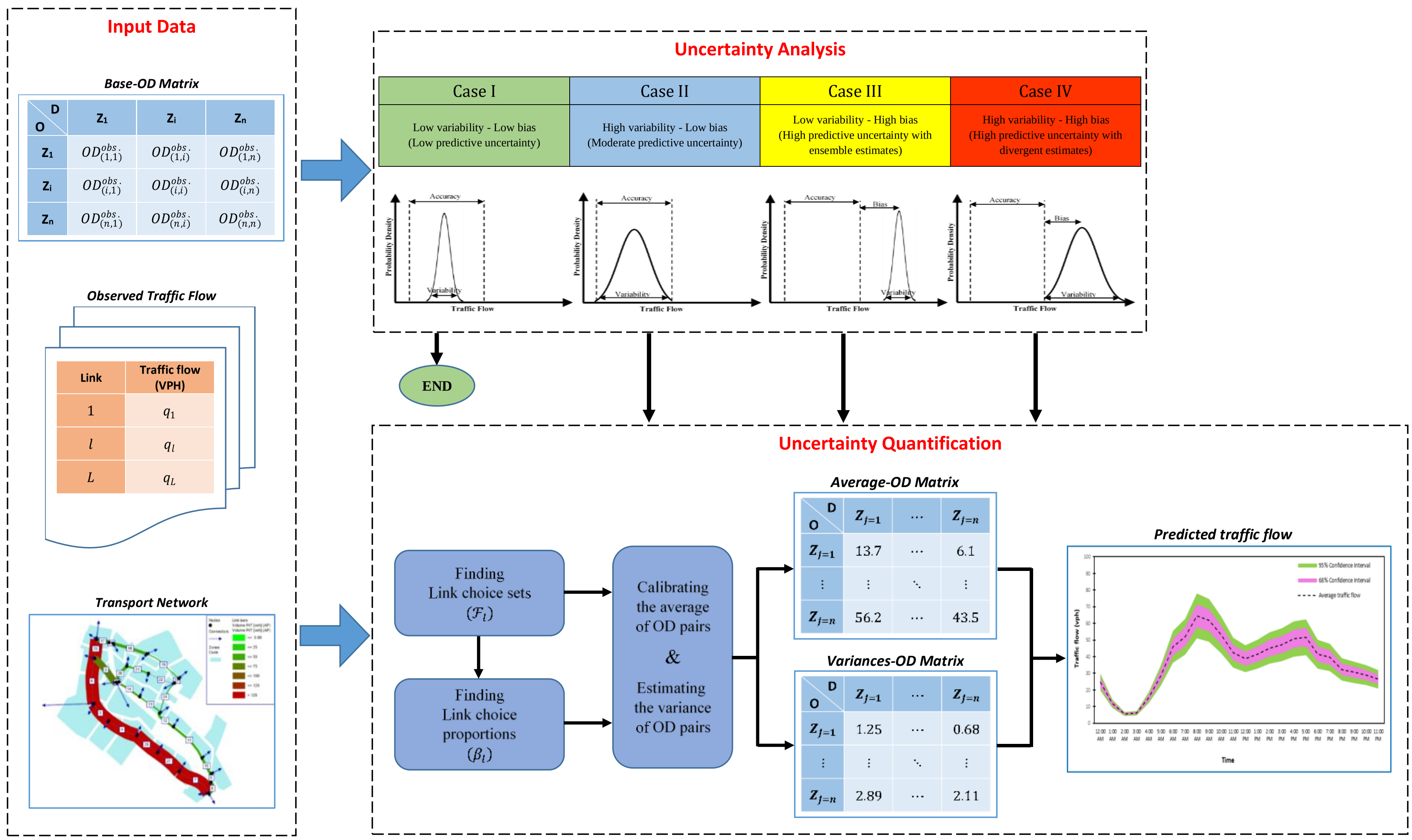


Figure 33: The uncertainty analysis and quantification framework.

6.4. List of Publications Related to the Theses

[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

References

[1] De Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., Hofman, F., "Uncertainty in traffic forecasts: literature review and new results for The Netherlands", Transportation, *34*(4), pp. 375–395, 2007. DOI:10.1007/s11116-006-9110-8

[2] Zhao, Y., Kockelman, K. M., "The propagation of uncertainty through travel demand models: An exploratory analysis", The Annals of Regional Science, *36*(1), pp. 145–163, 2002. DOI:10.1007/s001680200072

[3] Yang, C., Chen, A., Xu, X., Wong, S. C., "Sensitivity-based Uncertainty Analysis of a Combined Travel Demand Model", Procedia - Social and Behavioral Sciences, *80*, pp. 395–415, 2013. DOI:10.1016/j.sbspro.2013.05.022

[4] Rasouli, S., *Uncertainty in Modeling Activity-Travel Demand in Complex Urban Systems*. Ph.D. Thesis, Technische Universiteit Eindhoven, 2016.

[5] Rasouli, S., Timmermans, H., "Uncertainty in travel demand forecasting models: literature review and research agenda", Transportation Letters: The International Journal of Transportation Research, *4*(1), pp. 55–73, 2012. DOI:10.3328/TL.2012.04.01.55-73

[6] Ortúzar, J. de D., Willumsen, L. G., *Modelling Transport*, Modelling Transport (4th Editio.). John Wiley and Sons, Ltd., Publication, 2011. DOI:10.1002/9781119993308

[7] Mahmassani, H. S., "Uncertainty in transportation systems evaluation: issues and approaches", Transportation Planning and Technology, *9*(1), pp. 1–12, 1984. DOI:10.1080/03081068408717264

[8] Ottomanelli, M., Wong, C. K., "Modelling uncertainty in traffic and transportation systems", Transportmetrica, *7*(1), pp. 1–3, 2011. DOI:10.1080/18128600903244636

[9] Cambridge Systematics Inc., Vanasse Hangen Brustlin Inc., Gallop Corporation, Bhat, C. R., Shapiro Transportation Consulting LLC, Martin Alexiou Bryson PLLC, *Travel Demand Forecasting: Parameters and Techniques*, National Cooperative Highway Research Program, 2012. DOI:10.17226/14665

[10] Briggs, W., *Uncertainty: The soul of modeling, probability and statistics*, (William Briggs, Ed.)Uncertainty: The Soul of Modeling, Probability and Statistics. Switzerland: Springer International Publishing, 2016. DOI:10.1007/978-3-319-39756-6

[11] De Jong, G., Pieters, M., Miller, S., Daly, A., Plasmeijer, R., Graafland, I., … Kroes, E., *Uncertainty in Traffic Forecasts Literature Review and New Results for the Netherlands*. Netherlands, 2005. Retrieved from https://www.rand.org/content/dam/rand/pubs/working_papers/2005/RAND_WR268.pdf

[12] J, De Jong, G., Daly, A., Pieters, M., Miller, S., Hofman, F., "Uncertainty in traffic forecasts: literature review and new resuls for the Netherlands", European journal of transport and infrastructure research (EJTIR), pp. 1–23, 2005.

[13] Matas, A., Raymond, J. L., Ruiz, A., "Traffic forecasts under uncertainty and capacity constraints", Transportation, *39*(1), pp. 1–17, 2012. DOI:10.1007/s11116-011-

9325-1

[14] Van Der Laan, D. J., De Jonge, E., Solcer, J., "Effect of displaying uncertainty in line and bar charts: Presentation and interpretation", IVAPP 2015 - 6th International Conference on Information Visualization Theory and Applications; VISIGRAPP, Proceedings, (January 2015), pp. 225–232, 2015. DOI:10.5220/0005300702250232

[15] Ziems, S. E., Sana, B., Plotz, J., Pendyala, R. M., "Stochastic Variability in Microsimulation Modeling Results and Convergence of Corridor-Level Characteristics", *7066*(August), 2010.

[16] Manzo, S., Nielsen, O. A., Prato, C. G., "Investigating uncertainty in transport modelling: a four-stage model case study", Latsis symposium, 2012. Retrieved from http://transp-or.epfl.ch/heart/2012/latsis2012_submission_146.pdf

[17] Manzo, S., Nielsen, O. A., Prato, C. G., "How uncertainty in input and parameters influences transport model: Output A four-stage model case-study", Transport Policy, *38*, pp. 64–72, 2015. DOI:10.1016/j.tranpol.2014.12.004

[18] Rasouli, S., Timmermans, H., "Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research", Travel Behaviour and Society, *1*, pp. 79–90, 2014. DOI:10.1016/j.tbs.2013.12.001

[19] Pradhan, A., Kockelman, K. M., "Uncertainty propagation in an integrated land use-transportation modeling framework - Output variation via UrbanSim", Travel Demand and Land Use 2002, (1805), pp. 128–135, 2002. DOI:10.3141/1805-15

[20] Rider, W. J., Kamm, J. R., Weirs, V. G., "Verification , Validation and Uncertainty Quantification Workflow in CASL", Sandia National Laboratories, (SAND2010-234P), pp. 1–27, 2010. Retrieved from https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/ValUQWorkflowInCASL-RiderFinal.pdf

[21] Walker, W. E., Harremo, P., Es, € E, Rotmans, J., Van Der Sluijs, J. P., Van Asselt, A., … Krauss, V., "Defining Uncertainty A Conceptual Basis for Uncertainty Management in Model-Based Decision Support", *4*(1), pp. 5–17, 2003. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.7495&rep=rep1&type=pdf

[22] Walker, W. E., Lempert, R. J., Kwakkel, J. H., "Deep Uncertainty", In Encyclopedia of Operations Research and Management Science (pp. 395–402), 2013. DOI:10.1007/978-1-4419-1153-7_1140

[23] Ayyub, B., Klir, G., *Uncertainty Modeling and Analysis in Engineering and the Sciences*. 6000 Broken Sound Parkway NW: Chapman & Hall/CRC, 2006. DOI:10.1201/9781420011456

[24] Cools, M., Karlis, D., Janssens, D., Kochan, B., Bellemans, T., Wets, G., "Quantifying Input Uncertainty in Traffic Assignment Models", In Proceedings of the 91st Annual Meeting of the Transportation Research Board. Transportation Research Board of the National Academies, 2012. Retrieved from https://orbi.uliege.be/bitstream/2268/134333/1/TRB_Quantifying_input_uncertainty_in_traffic_assignment_models_REVISED.pdf

[25] Waller, S., Schofer, J., Ziliaskopoulos, A., "Evaluation with Traffic Assignment Under Demand Uncertainty", Transportation Research Record, *1771*(1), pp. 69–74, 2001. DOI:10.3141/1771-09

[26] Calafiore, G., Ghaoui, L. El, "Robust Dynamic Traffic Assignment under Demand and Capacity Uncertainty ∗", Uctc, (15833), pp. 1–10, 2008. Retrieved from https://pdfs.semanticscholar.org/eead/668e19889e025b9bb538407cf36dddd157b1.pdf

[27] Du, Y., Geng, Y., Sun, L., "Simulation model based on Monte Carlo method for traffic assignment in local area road network", Frontiers of Architecture and Civil Engineering in China, *3*(2), pp. 195–203, 2009. DOI:10.1007/s11709-009-0032-3

[28] Janssen, H., "Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence", Reliability Engineering and System Safety, *109*, pp. 123–132, 2013. DOI:10.1016/j.ress.2012.08.003

[29] Pang, A. T., Wittenbrink, C. M., Lodha, S. K., "Approaches to uncertainty visualization", The Visual Computer, *13*(8), pp. 370–390, 1997. DOI:10.1007/s003710050111

[30] Griethe, H., Schumann, H., "The Visualization of Uncertain Data: Methods and Problems", Proceedings of SimVis '06, *vi*(August), pp. 143–156, 2006. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Visualization+of+Uncertain+Data:+Methods+and+Problems#0

[31] Gosink, L., Bensema, K., Pulsipher, T., Obermaier, H., Henry, M., Childs, H., Joy, K., "Characterizing and visualizing predictive uncertainty in numerical ensembles through bayesian model averaging", IEEE Transactions on Visualization and Computer Graphics, *19*(12), pp. 2703–2712, 2013. DOI:10.1109/TVCG.2013.138

[32] Chang-Sung Jeong, Pang, A., "Reconfigurable disc trees for visualizing large hierarchical information space", In Proceedings IEEE Symposium on Information Visualization (Cat. No.98TB100258) (pp. 19-25,), 1998. DOI:10.1109/INFVIS.1998.729555

[33] Bonneau, G.-P., Hege, H.-C., Johnson, C. R., Oliveira, M. M., Potter, K., Rheingans, P., Schultz, T., "Overview and State-of-the-Art of Uncertainty Visualization", In Overview and State-of-the-Art of Uncertainty Visualization (pp. 3–27). Springer, London, 2014. Retrieved from http://www.sci.utah.edu/publications/Bon2014a/Overview-Uncertainty-Visualization-2015.pdf

[34] Iooss, B., Lemaître, P., "A Review on Global Sensitivity Analysis Methods", In Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications (pp. 101–122), 2015. DOI:10.1007/978-1-4899-7547-8_5

[35] Sobol, I. M., Kucherenko, S. S., "Global sensitivity indices for nonlinear mathematical models. Review", Wilmott, *2005*(1), pp. 56–61, 2005. DOI:10.1002/wilm.42820050114

[36] Cannavó, F., "Sensitivity analysis for volcanic source modeling quality assessment and model selection", Computers & Geosciences, *44*, pp. 52–59, 2012. DOI:10.1016/j.cageo.2012.03.008

[37] Borgonovo, E., Plischke, E., "Sensitivity analysis: A review of recent advances", European Journal of Operational Research, 2016. DOI:10.1016/j.ejor.2015.06.032

[38] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., … Tarantola, S., *Global Sensitivity Analysis. The Primer*, Global Sensitivity Analysis. The Primer, 2007. DOI:10.1002/9780470725184

[39] Pannell, D., "Sensitivity analysis of normative economic models: theoretical framework and practical strategies", Agricultural Economics, *16*(2), pp. 139–152, 1997. DOI:10.1016/S0169-5150(96)01217-0

[40] Borgonovo, E., Plischke, E., "Sensitivity analysis: A review of recent advances", European Journal of Operational Research. North-Holland, 2016, February 1. DOI:10.1016/j.ejor.2015.06.032

[41] Iooss, B., Lemaître, P., Iooss, B., Lemaître, P., "A review on global sensitivity analysis methods To cite this version : HAL Id : hal-00975701", 2015.

[42] Saltelli, A., Ratto, M., Saisana, M., Tarantola, S., *Global Sensitivity Analysis . The Primer*, Global Sensitivity Analysis. The Primer. John Wiley & Sons, 2008. DOI:10.1002/9780470725184

[43] Opgenoord, M. M. J., Allaire, D. L., Willcox, K. E., "Variance-Based Sensitivity Analysis to Support Simulation-Based Design Under Uncertainty", Journal of Mechanical Design, *138*(11), pp. 111410, 2016. DOI:10.1115/1.4034224

[44] Iooss, B., Saltelli, A., "Introduction : Sensitivity Analysis", In Springer Handbook on Uncertainty Quantification (pp. 1–31), 2015.

[45] Harenberg, D., Marelli, S., Sudret, B., Winschel, V., "Uncertainty Quantification and Global Sensitivity Analysis for Economic Models", SSRN Electronic Journal, 2017. DOI:10.2139/ssrn.2903994

[46] Borgonovo, E., Plischke, E., "Sensitivity analysis : A review of recent advances", European Journal of Operational Research, *248*(3), pp. 869–887, 2016. DOI:10.1016/j.ejor.2015.06.032

[47] Leurent, F., "Sensitivity and error analysis of the dual criteria traffic assignment model", Transportation Research Part B: Methodological, *32*(3), pp. 189–204, 1998. DOI:10.1016/S0191-2615(97)00024-6

[48] Ceylan, H., Bell, M. G. H., "Sensitivity analysis on stochastic equilibrium transportation networks using genetic algorithm", Journal of Advanced Transportation, *38*(3), pp. 291–321, 2004. DOI:10.1002/atr.5670380304

[49] Patriksson, M., "Sensitivity Analysis of Traffic Equilibria", Transportation Science, *38*(3), pp. 258–281, 2004. DOI:10.1287/trsc.1030.0043

[50] Shao, H., Lam, W. H. K., Tam, M. L., "A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand", Networks and Spatial Economics, *6*(3–4), pp. 173–204, 2006. DOI:10.1007/s11067-006-9279-6

[51] Kattan, L., Abdulhai, B., "Sensitivity analysis of an evolutionary-based time-dependent origin/destination estimation framework", IEEE Transactions on Intelligent Transportation Systems, *13*(3), pp. 1442–1453, 2012. DOI:10.1109/TITS.2012.2189437

[52] Sun, H., Gao, Z., Zhao, F., "Dynamic Network Design Problem under Demand Uncertainty: An Adjustable Robust Optimization Approach", Discrete Dynamics in Nature and Society, *2014*, pp. 1–16, 2014. DOI:10.1155/2014/739656

[53] Djukic, T., *Dynamic OD demand estimation and prediction for dynamic traffic management*. Delft University of Technology, 2014. Retrieved from http://repository.tudelft.nl/view/ir/uuid:ab12d7a7-e77b-424d-b478-d58657f94dd1/

[54] Ji, K., Ma, J., Tang, W., "Sensitivity analysis for stochastic user equilibrium traffic assignment with constraints", Special Issue Article Advances in Mechanical Engineering, *9*(5), pp. 1–7, 2017. DOI:10.1177/1687814017704354

[55] Chen, W., Jin, R., Sudjianto, A., "Analytical Variance-Based Global Sensitivity Analysis in Simulation-Based Design Under Uncertainty", Journal of Mechanical Design, *127*(5), pp. 875, 2005. DOI:10.1115/1.1904642

[56] Volostnikov, V. G., "Sensitivity Estimates for Nonlinear Mathematical Models", Computer Optics, *39*(4), pp. 459–461, 2015. Retrieved from https://pdfs.semanticscholar.org/d339/b9cc42d6a7286d96814e6713fd13cdde87e7.pdf

[57] Homma, T., Saltelli, A., "Importance measures in global sensitivity analysis of nonlinear models", Reliability Engineering & System Safety, *52*, pp. 1–17, 1996. DOI:10.1016/0951-8320(96)00002-6

[58] Bratley, P., Fox, B. L., "Algorithm 659: implementing Sobol's quasirandom sequence generator", ACM Transactions on Mathematical Software, *14*(1), pp. 88–100, 1988. DOI:10.1145/42288.214372

[59] Sobol′, I. ., "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates", Mathematics and Computers in Simulation, *55*(1–3), pp. 271–280, 2001. DOI:10.1016/S0378-4754(00)00270-6

[60] Jank, W., "Quasi-Monte Carlo sampling to improve the eeciency of Monte Carlo EM", Computational Statistics & Data Analysis, *48*, pp. 685–701, 2005. DOI:10.1016/j.csda.2004.03.019

[61] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index", Computer Physics Communications, *181*(2), pp. 259–270, 2010. DOI:10.1016/j.cpc.2009.09.018

[62] Ortuzar, J. de D., Willumsen, L. G., *Modelling Transport*, Modelling Transport, 2011. DOI:10.1002/9781119993308

[63] Avineri, E., Prashker, J., "Violations of Expected Utility Theory in Route-Choice Stated Preferences: Certainty Effect and Inflation of Small Probabilities", Transportation Research Record: Journal of the Transportation Research Board, *1894*, pp. 222–229, 2004. DOI:10.3141/1894-23

[64] Prato, C. G., "Route choice modeling: Past, present and future research directions", Journal of Choice Modelling, *2*(1), pp. 65–100, 2009. DOI:10.1016/S1755-5345(13)70005-8

[65] Sikka, N., "Understanding travelers' route choice behavior under uncertainty", Ph.D. Thesis, University of Iowa, pp. 166, 2012. DOI:10.17077/etd.49ibdit5

[66] Telgen, M. G., *Realistic route choice modeling*, Master thesis, University of Twente, 2010. Retrieved from https://essay.utwente.nl/59718/1/MA_thesis_M_Telgen.pdf

[67] Ton, D., Duives, D., Cats, O., Hoogendoorn, S., "Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam", Travel Behaviour and Society, *13*, pp. 105–117, 2018. DOI:10.1016/j.tbs.2018.07.001

[68] Ramming, M. S., "Network knowledge and route choice", Ph.D. Thesis, Massachusetts Institute of Technology, pp. 394 p., 2002. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.457.5449&rep=rep1&type=pdf

[69] Hoogendoorn-Lanser, "Modelling Travel Behaviour in Multi-model networks", Ph.D. Thesis, Technische Universiteit Delft, 2005. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:3e013e1d-5bd6-40d6-b1d5-87fc53f3b3d9?collection=research

[70] Bekhor, S., Ben-Akiva, M. E., Ramming, M. S., "Evaluation of choice set generation algorithms for route choice models", In Annals of Operations Research (Vol. 144, pp. 235–247), 2006. DOI:10.1007/s10479-006-0009-8

[71] Hoogendoorn-Lanser, S., Nes, R. Van, *ON THE USE OF CHOICE SETS FOR ESTIMATION AND PREDICTION IN ROUTE CHOICE*, Transport & Planning, Technische Universiteit Delft. Amsterdam, 2006. Retrieved from http://www.academia.edu/18439801/ON_THE_USE_OF_CHOICE_SETS_FOR_ESTIMATION_AND_PREDICTION_IN_ROUTE_CHOICE

[72] Bovy, P. H. L., Fiorenzo-Catalano, S., "Stochastic route choice set generation: Behavioral and probabilistic foundations", Transportmetrica, *3*(3), pp. 173–189, 2007. DOI:10.1080/18128600708685672

[73] Prato, C., Bekhor, S., "Applying Branch-and-Bound Technique to Route Choice Set Generation", Transportation Research Record: Journal of the Transportation Research Board, *1985*(1), pp. 19–28, 2007. DOI:10.3141/1985-03

[74] Van Nes, R., Hoogendoorn-Lanser, S., Koppelman, F. S., "Using choice sets for estimation and prediction in route choice", Transportmetrica, *4*(2), pp. 83–96, 2008. DOI:10.1080/18128600808685686

[75] Trucano, T. G., Swiler, L. P., Igusa, T., Oberkampf, W. L., Pilch, M., "Calibration, validation, and sensitivity analysis: What's what", Reliability Engineering and System Safety, *91*(10–11), pp. 1331–1357, 2006. DOI:10.1016/j.ress.2005.11.031

[76] Barnard, G. A., "Studies in the history of probability and statistics: ix. Thomas bayes's essay towards solving a problem in the doctrine of chances", Biometrika, *45*(3–4), pp. 293–295, 1958. DOI:10.1093/biomet/45.3-4.293

[77] Myung, I. J., "Tutorial on maximum likelihood estimation", Journal of Mathematical Psychology, *47*(1), pp. 90–100, 2003. DOI:10.1016/S0022-2496(02)00028-7

[78] Bovy, P., Stern, E., *Route Choice: Wayfinding in Transport Networks*. Kluwer Academic Publishers, 1990. DOI:10.1007/978-94-009-0633-4

[79] Frejinger, E., Bierlaire, M., Ben-Akiva, M., "Sampling of alternatives for route choice modeling", Transportation Research Part B: Methodological, *43*(10), pp. 984–994, 2009. DOI:10.1016/j.trb.2009.03.001

[80] Fiorenzo Calatano Stella, M., "Choice Set Generation in Multi-Modal Transportation Networks", Ph.D. Thesis, Delft University of Technology, pp. 333p, 2007.

[81] Mai, A. T., Bastin, F., Toulouse, M., *On Optimization Algorithms for Maximum Likelihood Estimation*, Interuniversity Research Center on Enterprise Networks, Logistics and Transportation, 2014. Retrieved from https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2014-64.pdf

[82] Jain, V., Doshi, P., "Model-Free IRL using Maximum Likelihood Estimation", 2015. Retrieved from www.aaai.org

[83] Seeger, M., "Gaussian Processes for Machine Learning", International Journal of Neural Systems, *14*(02), pp. 69–106, 2004. DOI:10.1142/S0129065704001899

[84] Andrianakis, Y., Challenor, P., "Parameter estimation and prediction using Gaussian Processes", pp. 1–16, 2009. Retrieved from http://mucm2.group.shef.ac.uk/Pages/Downloads/Technical Reports/09-05 YA 3.2.3 Parameter estimation and prediction using Gaussian.pdf

[85] Snelson, E. L., "Flexible and efficient Gaussian process models for machine learning", Ph.D. Thesis, University of London., pp. 1–135, 2007. DOI:10.1016/S1671-2927(09)60141-3

[86] Do, C. B., Lee, H., "Gaussian processes", Stanford University, Stanford, CA, pp. 1–14, 2007. Retrieved from http://cs229.stanford.edu/section/cs229-gaussian_processes.pdf

[87] Cascetta, E., "Travel-Demand Models", In Transportation Systems Analysis (pp. 169–257), 2009. DOI:10.1007/978-0-387-75857-2_4

[88] Krishnamurthy, S., Kockelman, K. M., Boothe, C., Professor, L., "Propagation of Uncertainty in Transportation-Land Use Models: An Investigation of DRAM-EMPAL and UTPP Predictions in Austin, Texas", Transportation Research Record, *1831*, pp. 219–229, 2003. DOI:10.3141/1831-25

[89] Seger, M., Kisgyörgy, L., "Predicting and Visualizing the Uncertainty Propagations in Traffic Assignments Model Using Monte Carlo Simulation Method", Journal of Advanced Transportation, pp. 1–11, 2018. DOI:10.1155/2018/9825327

[90] Deng, W., Lei, H., Zhou, X., "Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach", Transportation Research Part B: Methodological, *57*, pp. 132–157, 2013. DOI:10.1016/j.trb.2013.08.015

[91] Tennøe, S., Halnes, G., Einevoll, G. T., "Uncertainpy: A Python Toolbox for Uncertainty Quantification and Sensitivity Analysis in Computational Neuroscience", Frontiers in Neuroinformatics, *12*, pp. 49, 2018. DOI:10.3389/fninf.2018.00049

[92] Le Maître, O. P., Knio, O. M., *Spectral Methods for Uncertainty Quantification*. Dordrecht: Springer Netherlands, 2010. DOI:10.1007/978-90-481-3520-2

[93] Peter, A., *Uncertainty Quantification with Applications to Engineering Problems*. Technical University of Denmark, 2015. Retrieved from www.compute.dtu.dk

[94] Litvinenko, A., Matthies, H. G., "Inverse problems and uncertainty quantification",

2013. Retrieved from http://homepages.warwick.

[95] Wang, C., Wu, X., Kozlowski, T., "Gaussian Process–Based Inverse Uncertainty Quantification for TRACE Physical Model Parameters Using Steady-State PSBT Benchmark", Nuclear Science and Engineering, *193*(1–2), pp. 100–114, 2019. DOI:10.1080/00295639.2018.1499279

[96] Wu, X., Kozlowski, T., "Inverse uncertainty quantification of reactor simulations under the Bayesian framework using surrogate models constructed by polynomial chaos expansion", Nuclear Engineering and Design, *313*, pp. 29–52, 2017. DOI:10.1016/j.nucengdes.2016.11.032

[97] Ghanem, R., Owhadi, H., Higdon, D., *Handbook of uncertainty quantification*, Handbook of Uncertainty Quantification, 2017. DOI:10.1007/978-3-319-12385-1

[98] Wu, X., Shirvan, K., Kozlowski, T., "Demonstration of the Relationship between Sensitivity and Identifiability for Inverse Uncertainty Quantification", Journal of Computational Physics, 2019. DOI:10.1016/j.jcp.2019.06.032

[99] Cannavó, F., "Sensitivity analysis for volcanic source modeling quality assessment and model selection", Computers and Geosciences, *44*, pp. 52–59, 2012. DOI:10.1016/j.cageo.2012.03.008

[100] Nagel, J. B., *Bayesian techniques for inverse uncertainty quantification*. ZÜRICH, 2017. Retrieved from https://doi.org/10.3929/ethz-a-010835772

[101] Ashley, D. J., "Uncertainty in the context of highway appraisal", Transportation, *9*(3), pp. 249–267, 1980. DOI:10.1007/BF00153867

[102] Lowe, S. R., MORRELL, D., COPLEY, G., "Uncertainty in Highway Appraisal: the Development of Systematic Sensitivity Testing", In Planning & Transport Res & Comp, Sum Ann Mtg, Proc (pp. 15–29). PTRC Education and Research Services on behalf of Planning and Transport Research and Computation (International) Co, 1982. Retrieved from http://worldcat.org/isbn/0860501116

[103] Kroes, E. P., "Some Issues in Traffic Forecasting for Privately Financed Infrastructure", In European Transport Conference 1996, 1996.

[104] Leurent, F., "An analysis of modelling error, with application to a traffic assignment model with continuously distributed values of time", In TRANSPORTATION PLANNING METHODS. PROCEEDINGS OF SEMINAR D HELD AT THE PTRC EUROPEAN TRANSPORT FORUM, BRUNEL UNIVERSITY, ENGLAND, 2-6 SEPTEMBER 1996. VOLUME P404-1, 1996.

[105] Jong, G. de, Ettema, D., Gunn, H., Cheung, F., Kleijn, H., "Preparatory research into updating values of travel time in The Netherlands", In Proceedings WCTR Conference, 1998.

[106] Perrakis, K., Karlis, D., Cools, M., Janssens, D., Vanhoof, K., Wets, G., "A Bayesian approach for modeling origin-destination matrices", Transportation Research Part A: Policy and Practice, *46*(1), pp. 200–212, 2012. DOI:10.1016/j.tra.2011.06.005

[107] Soetaert, K., Petzoldt, T., "Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package FME", Journal of Statistical Software, *33*(3), 2010. DOI:10.18637/jss.v033.i03

[108] Frey, H. C., Rhodes, D. S., *Quantitative Analysis of Variability and Uncertainty in*

Environmental Data and Models. Volume 1. Theory and Methodology Based Upon Bootstrap Simulation, AAAS/EPA Environmental Science and Engineering. Oak Ridge, TN (United States), 1999. DOI:10.2172/1178941

[109] Smith, R. C., *Uncertainty quantification: theory, implementation, and applications*. Philadelphia: SIAM, 2013. Retrieved from http://bookstore.siam.org/cs12/

[110] Zio, E., Pedroni, N., *Methods for representing uncertainty*. Toulouse, France: Foundation for an Industrial Safety Culture, 2013. Retrieved from https://www.foncsi.org/fr/publications/collections/cahiers-securite-industrielle/literature-review-uncertainty-representation/CSI-uncertainty-representation.pdf

[111] Sullivan, T. J., *Introduction to Uncertainty Quantification* (Vol. 63). Cham: Springer International Publishing, 2015. DOI:10.1007/978-3-319-23395-6

[112] Hammonds, J. S., Hoffman, F. O., Bartell, S. M., "An Introductory Guide to Uncertainty Analysis in Environmental and Health Risk Assessment", Energy, pp. 47, 1994. DOI:10.2172/10127301

[113] Lin, G., Engel, D. W., Eslinger, P. W., *Survey and Evaluate Uncertainty Quantification Methodologies*, U.S. Department of Energy. Richland, WA (United States), 2012. DOI:10.2172/1035732

[114] Enki Yoo, E.-H., Aldstadt, J., *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses by Wenzhong Shi*, Journal of Regional Science (Vol. 51), 2011. DOI:10.1111/j.1467-9787.2011.00740_14.x

[115] Hand, D. J., "Measurements and their Uncertainties: A Practical Guide to Modern Error Analysis by Ifan G. Hughes, Thomas P. A. Hase", International Statistical Review, *79*(2), pp. 280–280, 2011. DOI:10.1111/j.1751-5823.2011.00149_8.x

[116] Ghanem, R., Owhadi, H., Higdon, D., *Handbook of Uncertainty Quantification*, (R. Ghanem, D. Higdon, & H. Owhadi, Eds.)Handbook of Uncertainty Quantification. Cham: Springer International Publishing, 2017. DOI:10.1007/978-3-319-12385-1

[117] Wu, X., Kozlowski, T., Meidani, H., Shirvan, K., "Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian process, Part 1: Theory", Nuclear Engineering and Design, *335*, pp. 339–355, 2018. DOI:10.1016/j.nucengdes.2018.06.004

[118] Iglesias, M. A., Lin, K., Stuart, A. M., "Well-posed Bayesian geometric inverse problems arising in subsurface flow", Inverse Problems, *30*(11), pp. 114001, 2014. DOI:10.1088/0266-5611/30/11/114001

[119] Litvinenko, A., Matthies, H. G., "Inverse problems and uncertainty quantification", 2013. Retrieved from http://arxiv.org/abs/1312.5048

[120] Litvinenko, A., Matthies, H. G., "Inverse problems and uncertainty quantification *", 2014. Retrieved from https://arxiv.org/pdf/1312.5048v2.pdf

[121] Navratil, G., *Adjustment Computations*, International Journal of Geographical Information Science (Vol. 25). Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010. DOI:10.1002/9780470586266

[122] Athanasios Dermanis, Armin Grün, F. S., *Geomatic Method for the Analysis of Data in the Earth Sciences*, (A. Dermanis, A. Grün, & F. Sansò, Eds.) (Vol. 95). Berlin,

Heidelberg: Springer Berlin Heidelberg, 2000. DOI:10.1007/3-540-45597-3

[123] Wolberg, J., *Data Analysis Using the Method of Least Squares*, Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments. Berlin/Heidelberg: Springer-Verlag, 2006. DOI:10.1007/3-540-31720-1

[124] Ogundare, J. O., *Understanding Least Squares Estimation and Geomatics Data Analysis*, Understanding Least Squares Estimation and Geomatics Data Analysis. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2018. DOI:10.1002/9781119501459

[125] Fagin, R., Halpern, J. Y., "Uncertainty, Belief, and Probability", In International Joint Conference on Artificial Intelligence (pp. 1161-1167.). Detroit, MI, 1991.

[126] Raychaudhuri, S., "Introduction to monte carlo simulation", Simulation Conference, 2008. WSC 2008., pp. 91–100, 2008. DOI:10.1109/WSC.2008.4736059

[127] Jantschi, L., Bolboaca, S. D., "Distribution Fitting 2. Pearson-Fisher, Kolmogorov-Smirnov, Anderson- Darling, Wilks-Shapiro, Cramer-von-Misses and Jarque-Bera Statistics", Bulletin UASVM Horticulture, *66*(2), pp. 691–697, 2009.

Appendix (A): Network Characterizations

A-1: Link List

NO.	FROM NODE NO.	TO NODE NO.	TRANSPORT SYSTEM	NUMBER OF ANES	PRIVATE TRANSPORT CAPACITY	SPEED (KM/H)	VOLUME (VPH)
1	1	4	BIKE, BUS, CAR, WALK	1	6000	40.0	73
2	4	1	BIKE, BUS, CAR, WALK	1	6000	40.0	60
3	2	9	BIKE, BUS, CAR, WALK	1	6000	30.0	366
4	9	2	BIKE, BUS, CAR, WALK	1	6000	30.0	409
5	9	16	BIKE, BUS, CAR, WALK	1	6000	30.0	365
6	16	9	BIKE, BUS, CAR, WALK	1	6000	30.0	408
7	1	2	BIKE, BUS, CAR, WALK	1	9000	50.0	395
8	2	1	BIKE, BUS, CAR, WALK	1	9000	50.0	388
9	3	10	BIKE, BUS, CAR, WALK	1	9000	50.0	407
10	10	3	BIKE, BUS, CAR, WALK	1	9000	50.0	363
11	3	8	BIKE, BUS, CAR, WALK	1	6000	40.0	38
12	8	3	BIKE, BUS, CAR, WALK	1	6000	40.0	48
13	7	8	BIKE, BUS, CAR, WALK	1	1500	20.0	48
14	8	7	BIKE, BUS, CAR, WALK	1	1500	20.0	38
15	7	15	BIKE, BUS, CAR, WALK	1	5000	30.0	22
16	15	7	BIKE, BUS, CAR, WALK	1	5000	30.0	32
17	14	15	BIKE, BUS, CAR, WALK	1	5000	30.0	27
18	15	14	BIKE, BUS, CAR, WALK	1	5000	30.0	16
19	6	14	BIKE, BUS, CAR, WALK	1	5000	30.0	27
20	14	6	BIKE, BUS, CAR, WALK	1	5000	30.0	17
21	6	13	BIKE, BUS, CAR, WALK	1	5000	30.0	8
22	13	6	BIKE, BUS, CAR, WALK	1	5000	30.0	19
23	5	13	BIKE, BUS, CAR, WALK	1	5000	30.0	26
24	13	5	BIKE, BUS, CAR, WALK	1	5000	30.0	15
25	4	5	BIKE, BUS, CAR, WALK	1	1500	20.0	57
26	5	4	BIKE, BUS, CAR, WALK	1	1500	20.0	45
27	21	22	BIKE, BUS, CAR, WALK	1	1500	20.0	39
28	22	21	BIKE, BUS, CAR, WALK	1	1500	20.0	40
29	17	22	BIKE, BUS, CAR, WALK	1	7000	40.0	40
30	22	17	BIKE, BUS, CAR, WALK	1	7000	40.0	39
31	21	23	BIKE, BUS, CAR, WALK	1	6000	40.0	28
32	23	21	BIKE, BUS, CAR, WALK	1	6000	40.0	28
33	20	23	BIKE, BUS, CAR, WALK	1	6000	40.0	10
34	23	20	BIKE, BUS, CAR, WALK	1	6000	40.0	10
35	5	18	BIKE, BUS, CAR, WALK	1	6000	40.0	31
36	18	5	BIKE, BUS, CAR, WALK	1	6000	40.0	30
37	18	24	BIKE, BUS, CAR, WALK	1	6000	40.0	18
38	24	18	BIKE, BUS, CAR, WALK	1	6000	40.0	18
39	19	24	BIKE, BUS, CAR, WALK	1	6000	40.0	0
40	24	19	BIKE, BUS, CAR, WALK	1	6000	40.0	0
41	19	20	BIKE, BUS, CAR, WALK	1	6000	40.0	9
42	20	19	BIKE, BUS, CAR, WALK	1	6000	40.0	9
43	6	19	BIKE, BUS, CAR, WALK	1	6000	40.0	9
44	19	6	BIKE, BUS, CAR, WALK	1	6000	40.0	9
45	1	17	BIKE, BUS, CAR, WALK	1	9000	50.0	392
46	17	1	BIKE, BUS, CAR, WALK	1	9000	50.0	413
47	16	25	BIKE, BUS, CAR, WALK	1	9000	50.0	364
48	25	16	BIKE, BUS, CAR, WALK	1	9000	50.0	408
49	10	25	BIKE, BUS, CAR, WALK	1	9000	50.0	407
50	25	10	BIKE, BUS, CAR, WALK	1	9000	50.0	363

A-2: Connector List

NO.	FROM ZONE NO.	TO ZONE NO.	DIRECTION	TRANSPORT SYSTEM	VOLUME (VPH)
1	1	17	O	BIKE, CAR, WALK	422
2	1	17	D	BIKE, CAR, WALK	400
3	2	2	O	BIKE, CAR, WALK	22
4	2	2	D	BIKE, CAR, WALK	72
5	3	3	O	BIKE, CAR, WALK	422
6	3	3	D	BIKE, CAR, WALK	388
7	4	4	O	BIKE, CAR, WALK	14
8	4	4	D	BIKE, CAR, WALK	14
9	5	4	O	BIKE, CAR, WALK	1
10	5	4	D	BIKE, CAR, WALK	1
11	6	13	O	BIKE, CAR, WALK	13
12	6	13	D	BIKE, CAR, WALK	13
13	7	14	O	BIKE, CAR, WALK	8
14	7	14	D	BIKE, CAR, WALK	8
15	8	15	O	BIKE, CAR, WALK	11
16	8	15	D	BIKE, CAR, WALK	11
17	9	7	O	BIKE, CAR, WALK	16
18	9	7	D	BIKE, CAR, WALK	16
19	10	20	O	BIKE, CAR, WALK	12
20	10	20	D	BIKE, CAR, WALK	12
21	11	24	O	BIKE, CAR, WALK	18
22	11	24	D	BIKE, CAR, WALK	18
23	12	18	O	BIKE, CAR, WALK	12
24	12	18	D	BIKE, CAR, WALK	12
25	13	23	O	BIKE, CAR, WALK	11
26	13	23	D	BIKE, CAR, WALK	11
27	14	21	O	BIKE, CAR, WALK	11
28	14	21	D	BIKE, CAR, WALK	11
29	15	23	O	BIKE, CAR, WALK	7
30	15	23	D	BIKE, CAR, WALK	7
31	16	6	O	BIKE, CAR, WALK	6
32	16	6	D	BIKE, CAR, WALK	6
33	17	19	O	BIKE, CAR, WALK	7
34	17	19	D	BIKE, CAR, WALK	7
35	18	25	O	BIKE, CAR, WALK	7
36	18	25	D	BIKE, CAR, WALK	7
37	19	16	O	BIKE, CAR, WALK	12
38	19	16	D	BIKE, CAR, WALK	12
39	20	9	O	BIKE, CAR, WALK	12
40	20	9	D	BIKE, CAR, WALK	12
41	21	10	O	BIKE, CAR, WALK	7
42	21	10	D	BIKE, CAR, WALK	7
43	22	3	O	BIKE, CAR, WALK	6
44	22	3	D	BIKE, CAR, WALK	6
45	23	2	O	BIKE, CAR, WALK	27
46	23	2	D	BIKE, CAR, WALK	28
47	24	1	O	BIKE, CAR, WALK	19
48	24	1	D	BIKE, CAR, WALK	19
49	25	17	O	BIKE, CAR, WALK	14
50	25	17	D	BIKE, CAR, WALK	14

A-3: Node List

NO.	TRANSPORT SYSTEM	VOLUME (VPH)
1	BIKE, BUS, CAR, WALK	841
2	BIKE, BUS, CAR, WALK	706
3	BIKE, BUS, CAR, WALK	23
4	BIKE, BUS, CAR, WALK	102
5	BIKE, BUS, CAR, WALK	102
6	BIKE, BUS, CAR, WALK	38
7	BIKE, BUS, CAR, WALK	54
8	BIKE, BUS, CAR, WALK	87
9	BIKE, BUS, CAR, WALK	762
10	BIKE, BUS, CAR, WALK	763
11	BIKE, BUS, CAR, WALK	21
12	BIKE, BUS, CAR, WALK	36
13	BIKE, BUS, CAR, WALK	38
14	BIKE, BUS, CAR, WALK	760
15	BIKE, BUS, CAR, WALK	31
16	BIKE, BUS, CAR, WALK	37
17	BIKE, BUS, CAR, WALK	11
18	BIKE, BUS, CAR, WALK	7
19	BIKE, BUS, CAR, WALK	56
20	BIKE, BUS, CAR, WALK	79
21	BIKE, BUS, CAR, WALK	20
22	BIKE, BUS, CAR, WALK	0
23	BIKE, BUS, CAR, WALK	764
24	BIKE, BUS, CAR, WALK	841
25	BIKE, BUS, CAR, WALK	706

Appendix (B): Visum via COM Interface Programming

1. Visum

Visum is a software system that enables transport engineers to model all private and public transport systems in one single integrated model. It is complemented by the microscopic traffic simulation system Vissim. Using Visum, most basic data provided by transport information and planning systems can be managed consistently and maintained with the network editor. Unlike simple GIS systems, Visum supports complex relations within single or several transport systems to be preserved, permitting transport engineers to create a suitable transport model.

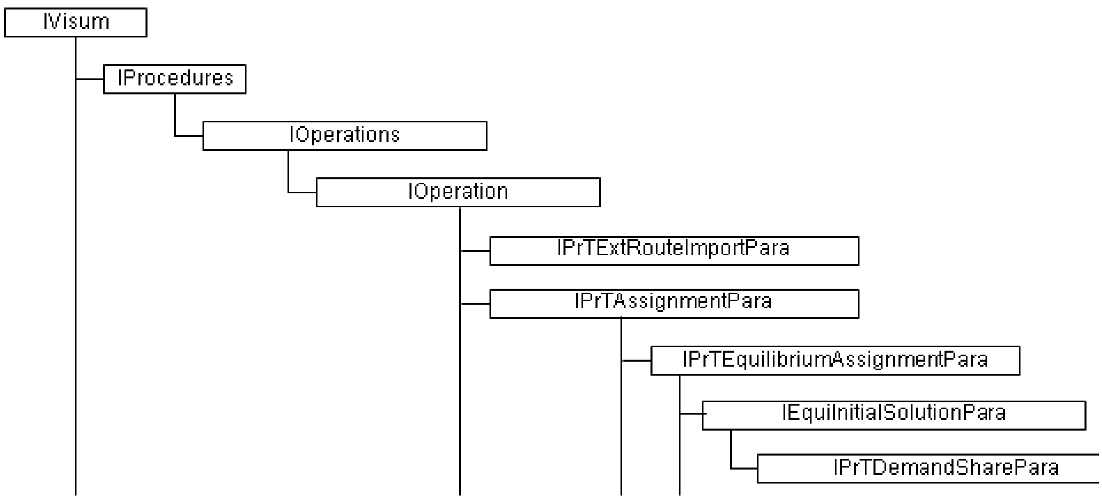
2. Visum COM interface

The Component Object Model COM explains how binary components of different software collaborate. COM gives access to data and functions contained in other programs. Data and algorithms contained in Visum can be accessed via the COM interface using Visum as an automation server. COM does not depend on a specific programming language; COM Objects can be included in a wide range of programming and scripting languages such as Visual Basic, Python, C++, Delphi, ..., and Matlab.

3. The basic steps of Visum-COM programming

The following procedure forms a general structure to use Visum-COM interface:

1. generating the Traffic Network (TN) and Traffic Analysis Zone (TAZ) through the Vissim GUI.
2. Choosing a programming language which allows COM interface programming. (Matlab was used in this study)
3. Programming the simulation via Visum-COM with specific commands. The Visum COM Model is subject to a strict object hierarchy (see Figure below). IVisum is the highest-ranking object. To access a sub-object, e.g. a link in the network, one must follow the hierarchy. The COM-interface reference included with the Online Help of Visum for details about the various objects and their methods and properties. (The letter "I" always represents the interface for the object).



4. Executing the simulation code form COM program.

4. Example

Complete example code for Visum-COM programming (written in Matlab) is presented below:

```
clc;
clear;
ME = 1;                                   % Mean value
SD = 0.05;                                % SD = Std. Dev. Value
N = 10000;                                % N = Nr. of Iterations
Z = 25;                                   % Z = Nr. of Zones
Visum = actxserver('Visum.Visum.17');
Visum.LoadVersion('Ajka.ver');
for k=1:N                                 % k = Iteration Nr.
  for i=1:Z
      for j=1:Z
       ODBase(i,j)= Visum.Net.ODPairs.ItemByKey(i,j).AttValue('MatValue(1M0101)');
      end
  end
 % Random Normal Distribution O-D Matrix
  for i=1:Z
    for j=1:Z
      ODIter(i,j)= (random('Normal',ME,SD)*ODBase(i,j));
    end
  end
  VectODIter = ODIter(:);                 % Save the O-D Matrix as a Vector
  TotalODIter(:,k) = VectODIter;          % Save all Vectors in One Matrix
 % Save New Random O-D Matrix in Visum
  for i=1:Z
    for j=1:Z
     Visum.Net.ODPairs.ItemByKey(i,j).set('AttValue','MatValue(1M0101)',ODIter(i,j));
    end
  end
  Visum.Procedures.Execute()
  LinkIter = Visum.Net.Links.Iterator;
  m = 1;                                  % m= link Nr.
  while LinkIter.Valid
       curLink = LinkIter.Item;
    Vol(m) = round(curLink.AttValue('VolVehPrT(AP)'));
    m = m+1;
    LinkIter.Next;
  end
  TotalVol(:,k)=Vol;                      % Save all Results in One Matrix
end
VolSD_05=TotalVol;
save VolSD_05.mat;
```