RIGA TECHNICAL UNIVERSITY
Faculty of Computer Science and Information Technology
Institute of Information Technology

Nadežda ZEŅINA
PhD Student of the Doctoral study program "Information Technology"

TRANSPORT TRAVEL DEMAND MODEL DEVELOPMENT BASED ON MACHINE LEARNING AND SIMULATION METHODS

Summary of the Doctoral Thesis

Scientific supervisor:
Professor *Dr. habil. sc. ing.*
J. MERKURJEVS

Scientific consultants:
Professor *Dr. habil. sc. comp.*
A. BORISOVS
Assoc. professor *Dr. sc. ing.*
A. ROMĀNOVS

RTU Press
Riga 2017

Zenina N. Transport travel demand model development based on machine learning and simulation methods. Summary of the Doctoral Thesis. – Riga: RTU Press, 2017. – 44 p.

Printed according to the Decision of the Institute of Information Technology as of January 13 2017, Minutes No. 12100-2/1.

This work was supported by the Latvian state research program project “The next generation of information and communication technologies (NexIT)”, (2014–2017).

ISBN 978-9934-10-942-3

DOCTORAL THESIS
PROPOSED TO RIGA TECHNICAL UNIVERSITY FOR THE PROMOTION TO THE SCIENTIFIC DEGREE OF DOCTOR OF ENGINEERING SCIENCES

To be granted the scientific degree of Doctor of Engineering Sciences, the present Doctoral Thesis will be publicly defended on June 12, 2017, at 14:30, Room 202, at the open meeting of the Faculty of Computer Science and Information Technology, Sētas Street 1.

OFFICIAL REVIEWERS

Professor *Dr. habil. sc. ing.* Zigurds Markovičs
Riga Technical University, Latvia

Professor *Dr.* Manfred Gronalt
BOKU – University of Natural Resources and Life Sciences, Austria

Dr. sc. ing. Mihails Savrasovs
Transport and Telecommunication Institute, Latvia

DECLARATION OF ACADEMIC INTEGRITY

I hereby declare that the Doctoral Thesis submitted for the review to Riga Technical University for the promotion to the scientific degree of Doctor of Engineering Sciences is my own and does not contain any unacknowledged material from any source. I confirm that this Thesis has not been submitted to any other university for the promotion to other scientific degree.

Nadežda Zeņina ……………………………. (Signature)

Date: ………………………

The Doctoral Thesis is written in Latvian and consists of an introduction, five sections, results analysis and conclusions, references, seven appendices, 55 figures, and 38 tables. The total volume of the Thesis is 163 pages. The bibliography contains 158 entries.

GENERAL THESIS DESCRIPTION

Introduction and Research Topicality

In today's rapidly changing urban environment, transportation engineers should be able to effectively estimate the planned transport infrastructure changes caused by the growth of transport flow intensities, and the new building construction and functional changes in the existing buildings. Transport travel demand model (TDM) is used to represent the selections of traveller trips. The representation of trip selection is influenced by different transport system characteristics such as traffic intensity, geometry of crossroads, the area and functionality of a building object, utility, social-economic indicators, etc. Transport travel demand models allow one to estimate future traffic changes after the changes in the functionality of a building object has been made and to identify the potential bottlenecks of transport system. Based on the model simulation results, transport planners can conclude about the effectiveness of transport solutions before they have been implemented. In spite of the variety of existing transport travel demand models [7], [20], [24], [80], [83], [93], [103], [104], [155], choosing a suitable transport travel demand model and methods to build the simulation model for transport impact analysis is still a challenging issue. The selected transport travel demand model should be easily transferable among territories, cost effective, and able to estimate the trip behaviour of separate travellers during the travel [24], [47], [126], [134].

The doctoral thesis contains a research on transport travel demand model development approach based on machine learning and simulation methods for transport impact analysis. Special attention in the simulation model is paid to the stages of transportation mode choice and generated transport trip number determination. The transport generated trip number is defined using the information systems that are based on regression equations. The known information systems contain the measurements about isolated territories that are not surrounded by pedestrian infrastructure and public transport; the building object has just one function, or a small amount of measurements is available. As a result, the transport generated trip number is measured either too high or too low. The present Doctoral Thesis provides an analysis of the usage of smart growth methods in the determination of transport generated trips aimed to ensure that generated transport trips are closer to the transport system under observation. However, when determining the selection of transportation mode in a trip-based transport travel demand model, the traveller's preferences in the selection of the transportation mode are not taken into account, because transportation mode choice is based on random utility maximization. It leads to a limited traveller trip behaviour analysis and does not fully provide the opportunity to evaluate transport infrastructure changes. In the research conducted within the Doctoral Thesis, the task of transportation mode choice has been completed using data mining methods and a decision tree algorithm that have flexible structures suitable for representing the relationships between the categories of transportation modes and do not require any knowledge about the initial specific model data structure.

The last stage of the transport travel demand model is model validation that includes verification and calibration. Transport travel demand model validation and calibration are usually done during its development so as to make sure that in the process of development in the particular stage the simulation results satisfy the initially defined terms and requirements. The Thesis summarizes the definitions of transport travel demand model validations, verifications and calibrations used by different researchers [26], [41], [54], [92] which point that there is no single unified definition for each meaning. In order to check whether the simulation model corresponds to the observed transport system, validation, calibration and verification are commonly made separately, though some researchers [26] include simulation model verification in the validation stage. The selection of simulation model calibration parameters and its value determination depends on the purpose of simulation model construction, as well as the complexity, quality and quantity of the gathered transport data and the type of the selected simulation model. In the Thesis, the task of global calibration parameter value selection was solved using clustering algorithms that can fast process a large amount of transport data and are able to work with uninterrupted and categorical variables when the amount of necessary clusters is unknown.

Research objective and tasks

The objective of the Doctoral Thesis was to elaborate the transport travel demand model development approach based on machine learning and simulation methods for transport impact analysis. To achieve the objective, the following research **tasks** have been set:

1) to investigate the types of transport travel demand models and their application in transport impact analysis;
2) to analyse the validation procedures of a transport travel demand simulation model (TDSM) for transport impact analysis;
3) to carry out an analysis of the usability of machine learning and simulation methods in transport demand models;
4) to define the machine learning and simulation methods for transport travel demand simulation model development;
5) to develop a technique for transportation mode choice based on a transport travel demand simulation model that takes into consideration traveller transportation mode preferences during the trip;
6) to develop a transport travel demand simulation model calibration procedure for calibration parameter value selection with clustering algorithms;
7) to design a transport travel demand simulation model for mixed-use building.

Research hypotheses

1. Transport smart growth methods improve the estimation of generated transport trips by analyzing the elements of transport system.
2. Transportation mode choice with data mining methods improves the classification results of transportation mode choices in transport travel demand simulation models and includes traveller transportation mode preferences during the trip.
3. Development of the transport travel demand simulation model calibration procedure provides a solution to the task of calibration parameter value selection with machine learning methods.

Research object and subject

The **research object** of the Doctoral Thesis is the transport travel demand simulation model. The **research subject** is machine learning and simulation methods.

Research methods used in the Doctoral Thesis

The research conducted in the Thesis includes an analysis of general and special literature, Latvian State laws, and Riga city regulation and statistical data analysis. The works of various foreign scientists, e. g., J. D. Ortúzar, L. G. Willumsen, M. E. Ben-Akiva, M. D. Meyer, E. J. Miller, etc., were used as a basis of theoretical researches for transport travel demand simulation model development. The machine learning, data mining and transport simulation methods were used in the present Thesis. To develop the transport travel demand simulation model, linear regression equations, smart growth methods, multinomial logit models, discriminant analysis, and Bi-level approach were employed. In the sample determination for transportation mode choice, minimum volume ellipsoid, minimum covariance determinant estimator, entropy and information gain as well as other data mining methods were used. When developing a technique for transportation mode choice, decision trees and statistical methods were used to evaluate the results obtained. During the development of transport travel demand simulation model calibration procedure, a combined simulation technology as well as simulation and clustering algorithms – SPSS Two-step clustering and hierarchical algorithms – were used.

The scientific novelty of the research

The scientific novelty of the Doctoral Thesis is the transport travel demand simulation model development approach, which allows estimating the results of transport infrastructure changes before they have been implemented and takes into account traveller transportation mode preferences during the trip. The elaborated transport travel demand simulation model development approach includes:

(1) smart growth methods analysis for generated transport trip determination;

(2) transportation mode choice technique for transport impact analysis;
(3) transport travel demand simulation model calibration procedure for transport impact analysis.

In the course of research, the intermediary results referred to the transport travel demand simulation model development approach were achieved: transport travel demand model types and their application in transport impact analysis were investigated; validation procedures of transport travel demand simulation model were analysed; combining of data mining methods was done for the selection of samples in order to decrease noisy data; initial gradient and traversal origin-destination matrix evaluation was done for trip assignment; and the transport travel demand simulation model for mixed-use building in Riga, in the area of Hanzas Street, was built.

Research practical value and approbation

The practical value of the Thesis is the transport travel demand simulation model development approach that enables simulation model development starting from conceptual model development to reliable assessment of modelled transport system measures for transport impact analysis.

The developed solutions have been proven in many transport impact analysis industrial projects. The author of the present Thesis has implemented these projects working as a leading transport infrastructure planning and modelling engineer for 10 years. Among them are "**Transport flow** research report for Shopping centre Alfa reconstruction (2015, Sia Solvers)", "**Transport flow** research and forecast project for multifunctional business centre "New Hanza City" at the territory among Brieža, Hanzas and Skanstes streets (2011, Sia Solvers)", "**Transport flow** and traffic infrastructure research, analysis and design project for Hanzas Crossing over the Daugava river (2010, Sia Solvers)", "**Transport flow** research and forecast project for the complex among Rūpniecības, Pētersalas and Katrīnas streets (2008, Sia Solvers)" determining the alternatives of transportation mode choices on promising developing objects by calibrating the transport travel demand simulation models and evaluating the alternatives of transport development with a simulation tool.

The intermediary results of transport travel demand simulation model development approach and developments were used in the Latvian State Research Program "The next generation of information and communication technologies (NexIT)" within the project "Sensor networks and signal processing applications" (2014–2017).

The results of the Doctoral Thesis are also used in the practical and laboratory classes of Riga Technical University within the subjects "Logistic information systems" and "Information technologies in logistics". Another use of the Thesis results in RTU is the lecture "Application of simulation in transport infrastructure development of the Riga city" delivered by the author to the students of IT Master's degree studies within the framework of the course "System simulation and modelling technology" since 2011.

Participation in international conferences

The results of the Doctoral Thesis were presented at and positive feedback was received from nine scientific conferences: RTU 57th International Scientific Conference (Latvia, 2016), RTU 56th International Scientific Conference (Latvia, 2015), The 17th International Conference on Harbor, Maritime & Multimodal Logistics Modelling and Simulation (Italy, 2015), The 8th International Conference on Urban Planning and Transportation (Spain, 2015), International Conference "Reliability and Statistics in Transportation and Communication (RelStat`14)" (Latvia, 2014), RTU 55th International Scientific Conference (Latvia, 2014), The 17th International Conference "Transport Means" (Lithuania, 2013), RTU 53rd International Scientific Conference (Latvia, 2011), and RTU 51st International Scientific Conference (Latvia, 2009).

Publications

Research findings were reported in 13 publications.

1. Zenina N., Romanovs A., Merkuryev J. Trip-based transport travel demand model for intelligent transport system measure evaluation based on micro simulation // International Journal of Simulation and Process Modelling, 2017 (accepted for publication).
2. Zenina N., Romanovs A., Merkuryev J. Transport simulation model calibration with two-step cluster // Scientific Journal of RTU, Computer Science. Information Technology and Management Science, Vol. 18, 2015, pp. 49–57. <EBSCO, CSA/ProQuest, VINITI>.
3. Zenina N., Romanovs A., Merkuryev J. Modelling Based Approach for Attracted Transport Readiness Trips Estimation to the Site // International Journal of Mathematical Models and Methods in Applied Sciences, Vol. 9, 2015, pp. 410–417. <SCOPUS>.
4. Merkuryev J., Zenina N., Romanovs A. Intelligent Transport Measures as a Component of Cyber-Physical Systems: Case Study for Adazi City // Proceedings of the 17th International Conference on Harbor, Maritime & Multimodal Logistics Modelling and Simulation, Italia, Bergeggi, Genova: DIME Universita di Genova, 2015, pp. 57–65. <SCOPUS>.
5. Zenina N., Romanovs A., Merkuryev J. Incoming Generated Traffic Flow Estimation based on Transport Access Design and Level of Service // In: Advances in Environmental Science and Energy Planning: Proceedings of the WSEAS 8th International Conference on Urban Planning and Transportation, Spain, Tenerife: WSEAS Press, 2015, pp. 195–201.
6. Zenina N., Merkuryev J. The impact of accessibility on transport infrastructure within commercial site // Reliability and Statistics in Transportation and Communication (RelStat`14), Latvia, Riga, 2014.
7. Zenina N., Borisovs A. Regression Analysis for Transport Trip Generation // Scientific Journal of RTU, Computer Science. Information Technology and Management Science, Vol. 16, 2013, pp. 89–94. <EBSCO, CSA/ProQuest, VINITI, Google Scholar>.
8. Zenina N., Borisovs A. Clustering algorithm for travel distance analysis // Scientific Journal of RTU, Computer Science. Information Technology and Management Science, Vol. 16, 2013, pp. 85–88. <EBSCO, CSA/ProQuest, VINITI, Google Scholar>.
9. Zenina N. Origin destination matrix estimation from traffic counts based on Bi level approach // Transport means 2013, Kaunas, Lithuania. <Web of Science (Conference Proceedings Citation Index), <SCOPUS>.

10. Zenina N., Borisovs A. Transportation Mode Choice Analysis Based on Classification Methods // Scientific Journal of RTU, Computer Science. Information Technology and Management Science, Vol. 16, 2011, pp. 49–53. <EBSCO, CSA/ProQuest, VINITI>.
11. Zenina N., Borisovs A. Artificial neural network for analysis of the traffic flow // Proceedings of 16th International Conference on Soft Computing – MENDEL 2010, Czech Republic, Brno, Brno University of Technology, 2010, pp. 25–25. <Thomson Reuters ISI Web of Science, SCOPUS>.
12. Zeņina N., Merkurjevs J. Analysis of simulation input and output to compare simulation tools // Mechanics Transport Communication, Bulgaria, Sofia, 2009, pp. V-4 – V-10.
13. Zenina N., Merkuryev J. An overview of artificial neural networks application in transportation // Proceedings of 14th International Conference on Soft Computing, Czech Republic, Brno, Brno University of Technology, 2008, pp. 2–15. <Thomson Reuters ISI Web of Science, SCOPUS >.

Outline of the Doctoral Thesis

The present Thesis consists of an introduction, five sections, a conclusion, bibliography, and seven appendices. The body of the Doctoral Thesis consists of 163 pages; it includes 55 pictures and 38 tables. Bibliographical references include 158 entries. The structure of the Thesis is described below.

Section 1 "**Transport demand models and their usability in transport impact analysis**" focuses on transport impact analysis. Types of transport travel demand models for transport impact analysis are examined, and transport travel demand model validation procedures are analyzed.

Section 2 "**Machine learning methods application in transport travel demand models in transport impact analysis**" deals with machine learning and simulation methods usability analysis in transport demand models. Methods used in the development of the transport travel demand simulation model are described.

Section 3 "**Transportation mode choice technique development for transport impact analysis**" proposes a technique for transportation mode choice with data mining and decision tree algorithms, which includes traveller transportation mode preferences during the trip. The section is dedicated to transport travel demand simulation model development.

Section 4 "**Transport travel demand simulation model calibration procedure development for transport impact analysis**" presents simulation model calibration procedure for the calibration parameter value selection with clustering algorithm. Calibration procedure provides the checking of the simulation model adequacy for the transport travel demand simulation model developed in Section 3.

Section 5 "**Transport travel demand simulation model development for mixed-use building**" describes the transport travel demand simulation model developed for mixed-use building in the area of Hanzas Street (Riga). The idea proposed in the present Doctoral Thesis was used in the simulation model building.

Results and Conclusions. References. Appendices.

SUMMARY OF THE THESIS SECTIONS

Section 1 "Transport demand models and their usability in transport impact analysis" contains an introduction to the research topic.

Transport impact analysis

Transport impact analysis [36], [126], [136] is performed to forecast, describe and analyse the future transport system behaviour after the proposed building objects are put into service to find out transport volume impact on the adjacent streets network, including related mitigation activities of the influence that would compensate it [45], [50], [55], [69], [81], [102]. A building object can be "a construction object", for example, a new building or "transport infrastructure object" – new intersection or access organization and existing intersection reconstruction.

Transport demand models

Transport demand model simulates the behaviour of a traveller during the trip, ensuring people movement from one point to another, carrying out some activities, using particular means of transport with the aim to forecast future transport demand [93], [104]. Transport demand models [7], [11], [22], [25], [31], [48], [52], [63], [64], [80], [82], [83], [103], [135], [155] can be expressed in different forms and ways (see Fig. 1).

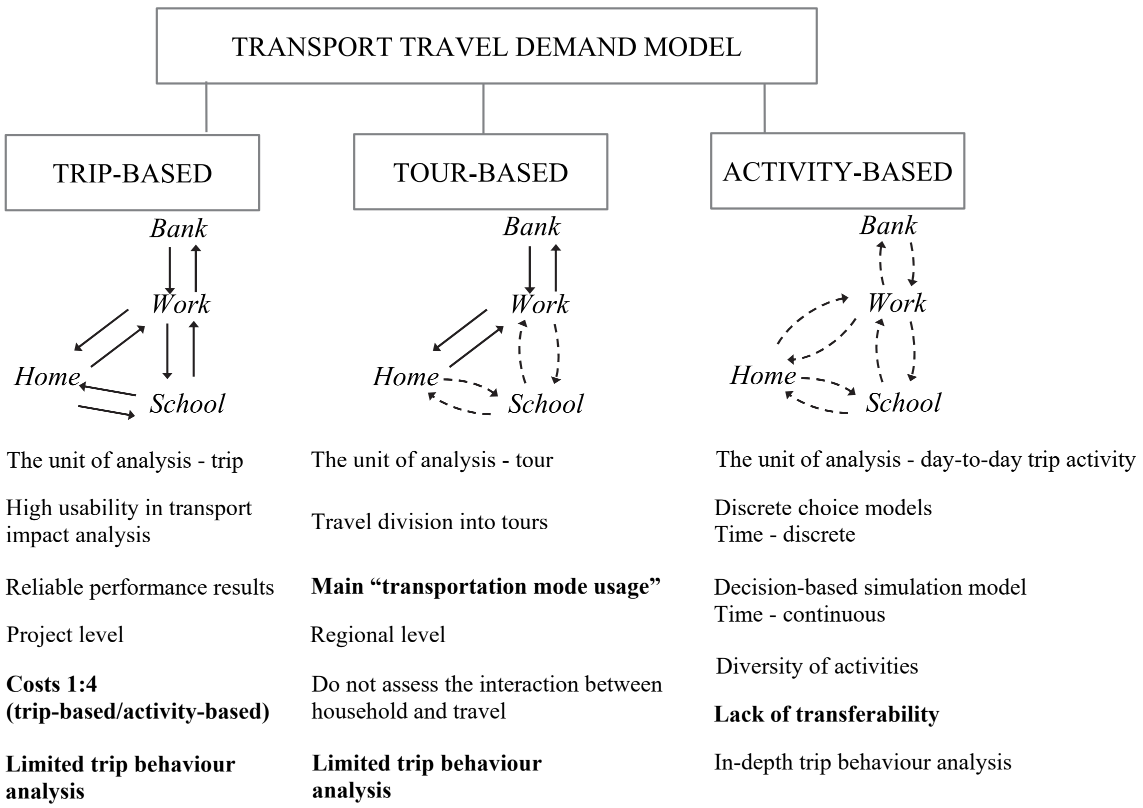


Fig. 1. Main characteristics of the transport travel demand model.

The **disadvantages** of the **trip-based** transport demand **model** (see Fig. 2) are related to the model property of focusing on aggregate behaviour rather than individual traveller trip behaviour and the **inability to assess the combined** transportation **modes** [103]. **Comparing** tour-based and trip-based transport travel demand **models** for different scenarios at regional and project levels, it was discovered that the trip-based transport travel demand model provides more acceptable transport intensity forecast results at a project level [47], [83]. **Activity-based transport travel demand models** in their turn are not widely used in practice [20] because model development can last for many years and require a large amount of data [24], [99], [134].

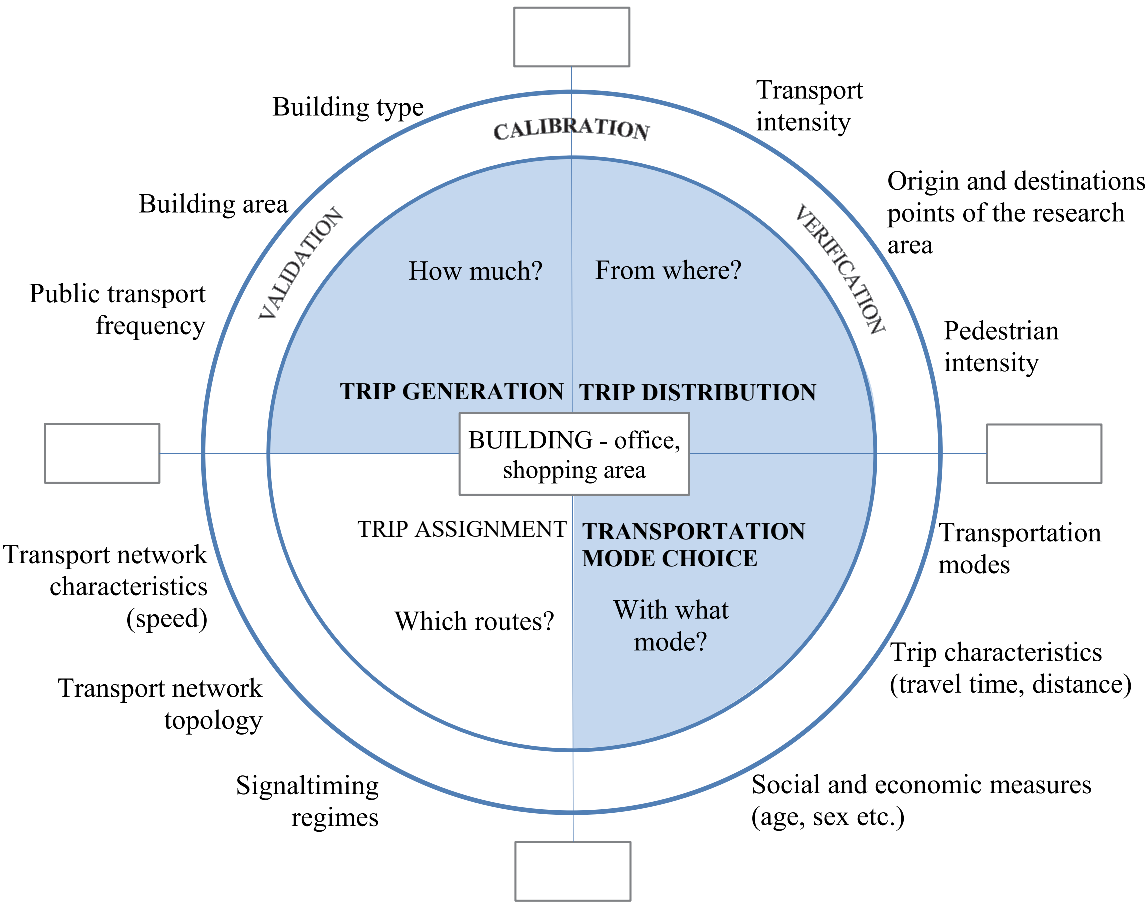


Fig. 2. Traditional trip-based transport travel demand model.

In the Thesis, the **trip-based** transport travel **demand model** is **selected for** simulation model **development** because the tour-based and activity-based transport travel demand models are difficult to transfer between development areas; besides, their development time and costs exceed several times the trip-based transport travel demand model development time and costs.

Transport travel demand model validation

Nowadays, simulation is becoming one of the common research methods in the field of transportation that allows comprehensive examination of the various processes and systems as well as evaluation of the expected results before decision making [40], [73], [92], [149]. The integrated transport simulation tool is used in the Thesis to develop and analyse the expanded trip-based transport travel demand model at the project level for transport impact analysis.

The validation and calibration of the transport travel demand simulation model (TDSM) is usually done during its development in order to make sure that in the process of development in the particular stage, the simulation results satisfy the initially defined terms and requirements [26], [127]. There is no single unified validation, verification and calibration definition [26], [41], [54], [92]. In the Thesis, transport travel demand **simulation model validation** is defined as a process in which it is checked how well transport travel demand simulation model fits the observed transport system. This is achieved by comparing the simulated and observed transport system logic and structure, by **adjusting TDSM calibration parameters** in the way that they are **consistent with the observed transport system**, and by checking if simulation model effectiveness measures adequately describe the observed transport situation. TDSM validation includes validation and calibration steps where simulation model verification occurs during the validation process. Validation methods can be divided into [26], [79], [91], [111]:

1) quality methods (e.g., animation, histogram) and
2) quantitative methods (e.g., percent error, root mean square error, mean percentage error, GEH statistics or Theil's uncertainty coefficient).

The selection of **validation** and **calibration parameters** and transport system effectiveness measures depends on the complexity of analysable task and the initial amount and quality of input data. In many cases, the selection of validation and calibration parameters can be a challenge because the determination of certain parameters requires specific observation methods that are not available to model users (e.g., detailed information about the driver's location and car's speed in the analyzed period and time with a certain interval). The developer of TDSM model must look for a compromise between the model development costs and the expected model detail level and accuracy [133]. Summarizing validation methods and parameters [13], [17], [35], [38], [72], [75], [76], [94], [95], [96], [98], [100], [105], [106], [110], [122], [133], [138], it can be concluded that the authors make model validation and calibration based on a model parameters' subset. They check the difference between the observed and simulation data in a certain period on a typical day and use different parameters of transport system effectiveness, or demand template is calibrated in advance. There are many TDSM calibration procedures available but according to the literature summary, more frequently used are those observed in [14], [19], [39], [41], [42], [57], [65], [79], [127]. In practice, to choose calibration parameters, an expert's opinion is most frequently used because other methods are too time-consuming [124].

The conclusions of Section 1 are as follows:

1) *the usage of models based on travels and activities for the objectives of the project-level traffic flow impact analysis is difficult because the development and costs of these models exceed several times the standard trip transport demand models costs and development time;*
2) *extended, trip-based transport demand model development for the analysis of traffic flow impact requires new methods and approaches that include traveller transportation mode preferences during the trip;*
3) *in the selection of calibration parameter values, methods are needed that could provide acceptable transport travel demand model accuracy and match the transport impact analysis requirements in terms of the execution time.*

Section 2 **"Machine learning methods application in transport travel demand models in transport impact analysis"** is dedicated to simulation and machine learning methods application analysis for transport travel demand simulation model development in transport impact analysis.

In each transport travel demand model stage, i.e., determination of generated trip number, transportation mode choice and trip assignment, a variety of mathematical, analytical and simulation models are used depending on the objective and tasks of transport impact analysis [60].

Cross-classification, neural network, linear regression equations [6], [27], [59], [89], [123] are used to **determine** the **number** of the **generated trips** [32]. **Cross-classification tables** are used in trip generation at strategic and tactical levels [66]; disadvantages of the method are widely described in the literature [33], [68], [118].

Neural networks are widely used as a data analysis technique in the transport trip determination [6], [59]. In the Thesis, multilayer perceptron with backpropagation (see Fig. 3) is utilized to solve the transport trip determination task [145] and to forecast transport trips seven days ahead.

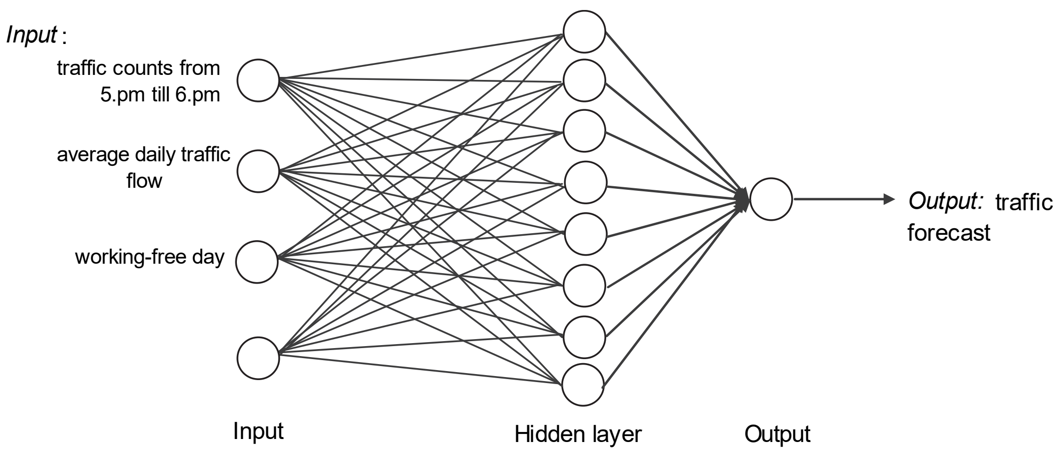


Fig. 3. Neural network learning algorithm [145].

The performed experiments showed that the accuracy of transport trip forecast generated by predictor varies within 15–40 % depending on the learning data set size. The evaluation of weight effect and the study of archetypes show that each connection has different influence on the result, both positive and negative. The analysis of the solution sensitivity to the changes in the value of one input variable showed positive influence of variables on the result. **Linear regression** method is the most frequently used method to evaluate the number of generated trips based on historical data in transport impact analysis [61], [93]. The main disadvantage of linear regression of trip generation is the necessity of making additional trip generation result corrections for adjacent sites and public transport availability near the site.

Despite some disadvantages, **linear regression methods** have some significant **advantages** [60] in the determination of the number of the generated trips:

1) mixed-used types and sizes of the land are taken into account when the number of generated trips is determined and the internal rate, which reduces the total number of generated trips, is calculated;
2) flexibility in the results is ensured by taking into account transport planning engineer considerations, which results in easy assign and interpretations of transport trip rates;
3) quality solutions are provided without using specialized and expensive software.

There are several **information systems** such as TRICS [86, 128], NZTPD and RTA [130], and ITE Trip Generation [60], [61] for trip generation, which are based on linear regression equations and average rates. In the Doctoral Thesis, for trip generation, the information system *ITE Trip Generation* was used with additional corrections made at the automobilization level for local transport conditions [114] taking into account that other information systems provide similar trip generation results or the information system does not have a sufficient number of researches for land use. With the information system *ITE Trip Generation*, the generated trips are **determined** mostly **for an isolated territory** without public transport accessibility and pedestrian infrastructure. To improve the accuracy of generated trips, it is proposed to use smart growth methods.

There are several **smart growth methods** (see Fig. 4) for trip generation correction in the ITE Trip Generation information system: Mixed-used, Urbemis, ERA MXD, NCHRP 8–51 [43], [61], [89], [90]. Most of smart growth methods focus on one trip generation reducing parameter, for example, land mixed-use. The smart growth method *Urbemis*, in turn, allows one to evaluate the developments' influence on various land-use types, public transport, parking places, pedestrian activities and other characteristics, despite the fact that the method was primarily developed for air quality evaluation.

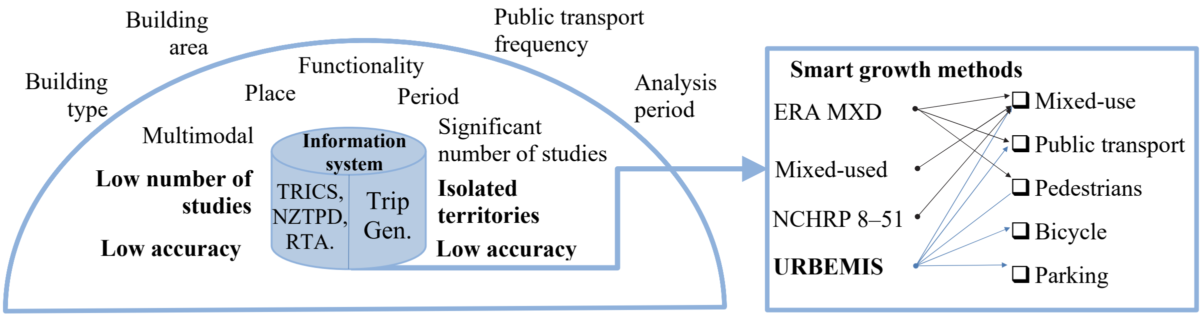


Fig. 4. Information systems and smart growth methods for trip generation.

Trip distribution is done by **converting the** origin and destination **transport intensity** including generated transport intensity **into origin-destination matrix.** For that purpose, simulation methods [60], such as Fratar method, gravity method [30], [119], network equilibrium method, entropy maximization and information minimization method [74], [116], [140], [141], [156], Bayesian inference method [78], least squares method [12], [23], [29], interference–opportunity method [1], [19], [49], [139], route distribution–assignment and bi-level approach [15], [21], [117] are used. Trip distribution methods can be classified by [34] transport network characteristics. In the Thesis, the bi-level approach is used for trip distribution for transport impact analysis, which provides origin-destination matrix determination in an over-congested street network taking into account the traveller's route selection.

As a result of transportation mode choice, the origin-destination matrix is developed, in which the number of travellers' transport trips from one zone to another by transport modes available in the research, is determined. Transportation mode choice includes sample selection and appropriate mode choice model selection in traveller trip behaviour simulation context [80]. **Sample selection** includes transport impact analysis target object determination, missing and noisy [5], [53] transport data analysis, and records selection [56], [70]. Traditional transportation **mode choice methods** and **models**, for example, multinomial logit model [18], [44], [93], [97], discriminant analysis [88], [120], [121], [157], [158], are based on random utility maximization principle and do not include the traveller's transportation mode preferences during the trip. **Summarizing** up-to-date **machine learning** methods [108], [112], [142] for transportation mode choice for various transport travel demand model types, it can be concluded that machine learning methods application in transportation mode choice is still a topical problem. Here it is taken into account that machine learning methods have a more flexible structure ratio between mode categories than the traditional logit models, and do not require prior knowledge about the initial specific model data structure. The application of the decision tree algorithms in tour-based and activity-based transport travel demand models has shown good performance results [62], [144]. The author of the Doctoral Thesis proposes to use the decision tree algorithms C4.5 and CART in trip-based transport travel demand model development [101].

As a result of trip assignment, transport trips have been assigned by transport routes. Dynamic assignment realized in integrated transport simulation software has been used in transport travel demand simulation model development, because it allows adjusting transport routes during the journey and analysing various transport alternatives with and without congestions [60], [131].

The conclusions of Section 2 are as follows:

1) *information systems used for the determination of the number of trips in trip-based transport travel demand model, do not take into account the presence of traffic participants in the analyzed territories, or the number of available measures is very small. Therefore, the issue of application of smart growth methods that would allow one to approximate the generated trips to the observed transport system is topical;*
2) *application of linear regression and smart growth methods in trip generation does not require the use of any specialized and expensive software;*
3) *transportation mode choice with traditional methods (multinomial logit, discriminant analysis) in the trip-based transport travel demand model does not include transportation mode preferences of the individual traveller, which leads to a limited individual traveller trip behaviour analysis and does not fully provide an opportunity to evaluate the transport travel demand model transport infrastructure changes before they have been implemented.*

Section 3 "**Transportation mode choice technique development for transport impact analysis**" is dedicated to transport travel demand simulation model and transportation mode choice technique development. The process of transport travel demand conceptual model development is shown in Fig. 5.

In the Thesis, transport travel demand simulation model development approach is based on simulation and machine learning methods [80], [112], [125], [142]. These methods allow one to adjust the determined generated trips to the observed transport system behaviour and to account the traveller transportation mode preferences during the trip in transportation mode choice by providing more accurate classification results in comparison with traditional methods for transportation mode choice.

Determination of transport generated trip number

To test the research hypothesis that transport smart growth methods improve the estimation of generated transport trips by analyzing the elements of transport system, the transport generated trip number for various building objects was determined (see Table 1).

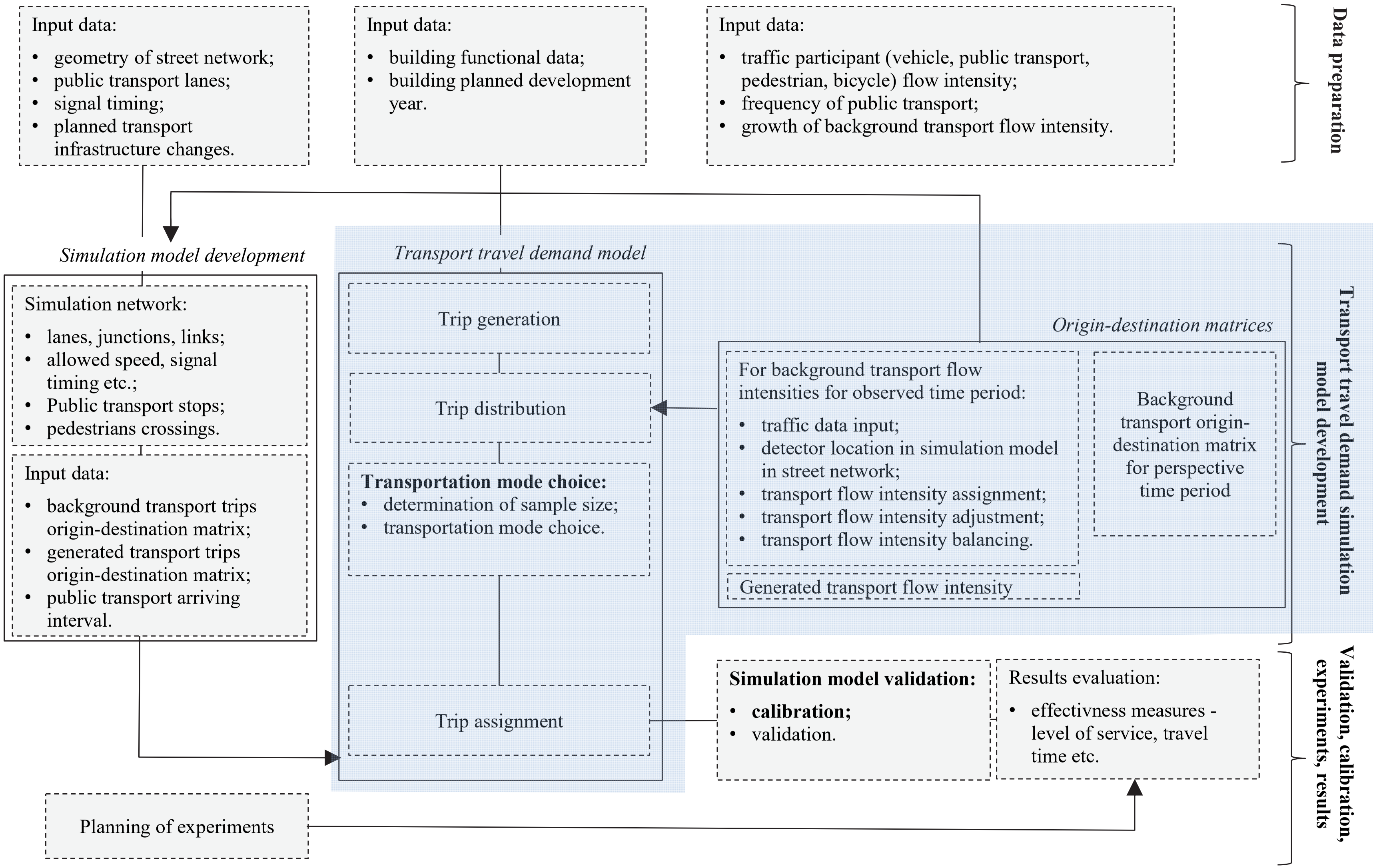


Fig. 5. Transport travel demand conceptual model for transport impact analysis.

Table 1

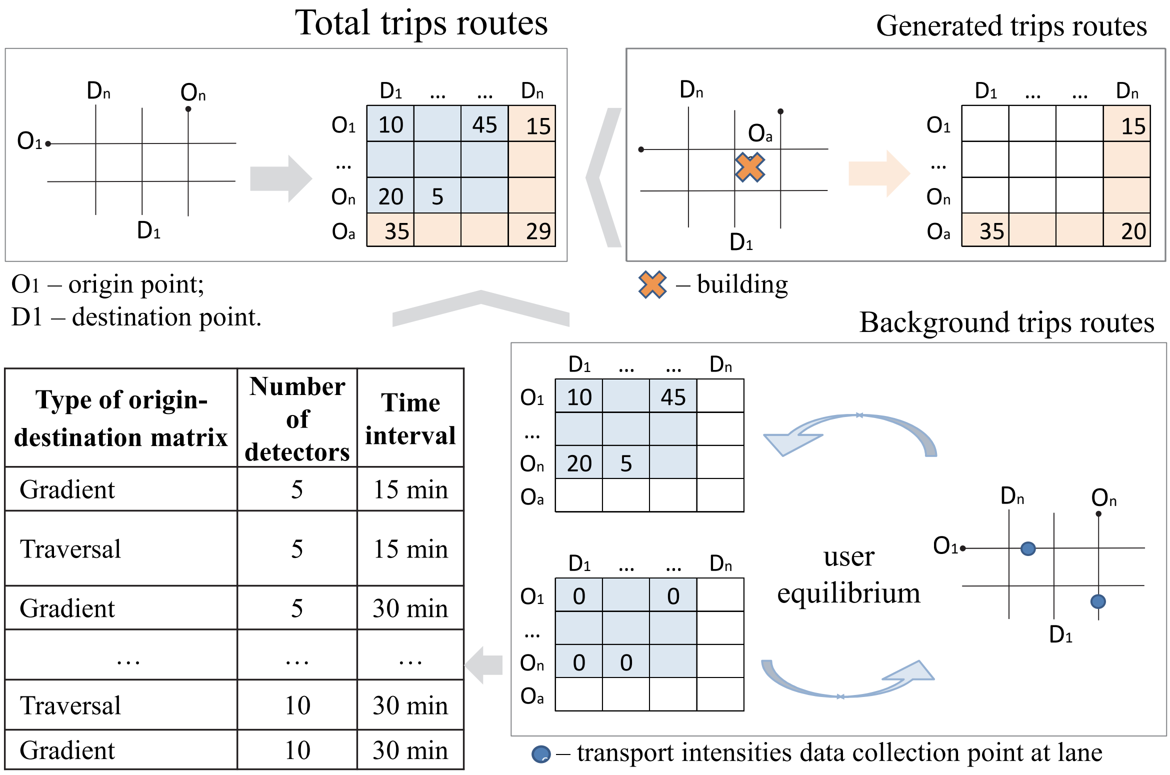
Percent error between the observed and actual transport generated trips for the evening peak hour

Mixed-use building	**Information system *ITE Trip Generation***	**Information system *ITE Trip Generation* with corrections**			**Information system *TRICS***	**Information system *NZTPD***
		Urbemis	***ERA MXD***	***NCHRP 8–51***		
1	34 %	31 %	17 %	30 %	**10 %**	22 %
1a	**0 %**	**3 %**	38 %	**3 %**	18 %	9 %
2	**5 %**	9 %	28 %	6 %	31 %	6 %
3	6 %	**0 %**	13 %	1 %	6 %	4 %
3a	10 %	**3 %**	13 %	7 %	10 %	23 %

Experimental results have shown that trip generation accuracy [26] does not depend only on building functional characteristics but also on building location in the street network, public transport accessibility around the building, and on pedestrians and bicycle infrastructure. Such transport infrastructure parameters as the number of employees in the research area, the number of pedestrians, land-use type and parking existence can significantly improve or worsen the determination accuracy of transport generated trip number. If the building is located near a highway without developed public transport network and pedestrians infrastructure, then the information system *ITE Trip Generation* can be used for trip generation in transport impact analysis. If the building adjacent territory has a well developed public transport network and the pedestrian and bicycle infrastructure, then the information system *ITE Trip Generation* with additional corrections for local conditions and with smart growth method *Urbemis* can be used for trip generation in transport impact analysis. The application of the smart growth method *Urbemis* in trip generation allowed adjusting the generated trips taking into account public transport, pedestrian and bicycling infrastructure in the research area.

Trip distribution

Trip distribution determination process is depicted in Fig. 6. The bi-level approach described in Section 2 is used for background transport flow intensities convertation to origin-destination matrices. To perform the origin and destination attraction point convertation, firstly, the initial matrix should be defined, from which the iterative adjustment process will begin. In the Thesis, two initial origin-destination matrices are analysed: gradient and traversal with various transport flow intensity time and number measurements [137].



Type of origin-destination matrix	Number of detectors	Time interval
Gradient	5	15 min
Traversal	5	15 min
Gradient	5	30 min
…	…	…
Traversal	10	30 min
Gradient	10	30 min

Fig. 6. Trip distribution determination for a building object in transport impact analysis.

Trip distribution results for various numbers of detectors and time periods have shown that origin-destination matrix transport flow intensity is slightly underestimated if the initial matrix is developed based on gradient method when the number of detectors is more than five. If the number of detectors in street network does not exceed five, transport flow intensity becomes slightly overestimated.

Transportation mode choice. Determination of the sample size

Transportation mode choice technique includes determination of sample size and transportation mode choice. The determination of sample size is the first step in the transportation mode choice technique. The diagram of sample size determination is shown in Fig. 7.

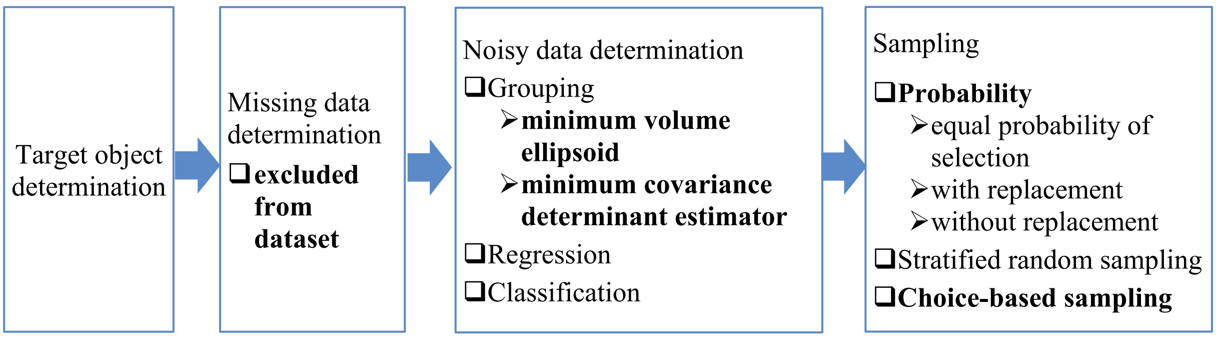


Fig. 7. Determination of sample size.

For determination of sample size, the dataset with 7171 respondents is available. The dataset includes the following attributes: traveller age, travel time, transportation mode, location from, location to, building land-use – shopping area, office area, respondents direction (into building, out of building), weather, and day of week.

For sample noisy data determination together with the minimum volume ellipsoid and minimum covariance determinant estimator, various data mining methods are used [148]. The application example of the minimum volume ellipsoid is shown in Fig. 8.

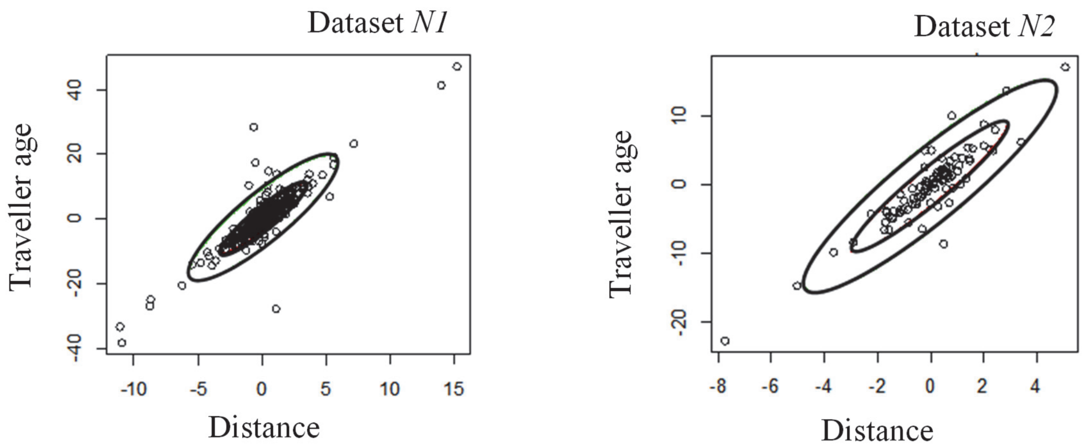


Fig. 8. Application example of the minimum volume ellipsoid for noisy data determination.

Datasets and transportation mode choice attribute percentage distribution are given in Table 2 obtained from sample determination results. Initial datasets records were processed taking into account information gain and cost-sensitive analysis [148] to improve the determination of the sample size and classification accuracy.

Table 2

Categories distributions for "Transportation mode choice" attribute

Attribute "Transportation mode choice" categories	Initial dataset *N0*	Dataset *N1*	Dataset *N2*	Dataset *N3*
Vehicle	47 %	22 %	25 %	34 %
Pedestrian	30 %	22 %	24 %	33 %
Bicyclist	1 %	17 %	25 %	*Excluded from dataset*
Public transport	20 %	22 %	26 %	33 %
Taxi	1 %	17 %	*Combined with vehicles*	*Combined with vehicles*
Total:	100 % (7 171 records)	100 % (498 records)	100 % (560 records)	100 % (1 380 records)

Record classification accuracy has improved insignificantly within 1–2 %, and for separate datasets it has worsened for 2–3 % by excluded attributes from dataset *N1* with information gain under 0.40 (three attributes met the condition "building land-use", "respondents direction", and "day of week"). The improvement or disimprovement of classification results for various datasets within 1–3 % are related to probability. The number of correctly classified instances has improved by five percent (see Table 3) after cost-sensitive matrix evaluation together with minimum volume ellipsoid was performed for datasets *N2* and *N3*.

Table 3

Percentage of correctly classified instances for datasets *N2* and *N3* via algorithm *C4.5 example*, by combining data mining methods

Methods for determination of sample size	**Initial dataset *N2***			**Initial dataset *N3***		
	Without 1 atribute[1)]	**Without 3 attributes[2)]**	**Cost-sensitive analysis**	**Without 1 atribute[1)]**	**Without 3 atributes[2)]**	**Cost-sensitive analysis**
Minimum volume ellipsoid	76 %	77 %	80 %	75 %	78 %	82 %
K-fold cross-validation	77 %	75 %	79 %	77 %	74 %	80 %

1) – Attribute "date" does not participate in classification;
2) – Attributes "day of week", "building land-use", and "respondents direction" do not participate in classification.

Transportation mode choice

In the Thesis, to test the second hypothesis, transportation mode choice was performed using decision tree algorithms described in Section 2.

Decision tree algorithms

C4.5 and CART decision tree algorithms were used for transportation mode choice in the process of transport travel demand simulation model development. Datasets *N1–N3* with and without the "taxi" and "bicycle" transportation mode choice attribute categories were used for instances classification.

In addition to datasets, the noise reduction approach complex was used (dataset *N3+*), combining minimum volume ellipsoid and random sampling to reduce the number of noisy data and to provide a more uniform distribution of transportation mode choice categories in the dataset. Classification results are given in Table 4.

The results obtained showed that the number of correctly classified instances was over 77 % for both decision tree algorithms. The classification results of transportation mode choice have improved by three percent for C4.5 decision tree algorithm and by five percent for CART decision three algorithms after attributes "building land-use" and "weather" were excluded from the analysis.

Table 4

Percentage of correctly classified instances for dataset *N3+*

Transportation mode choice method	Initial dataset *N3+* and combined data mining methods		
	Without *1* attribute[1)]	Without *2* attributes[2)]	Without *3* attributes[3)]
C4.5	77 %	80 %	79 %
CART	77 %	91.8 %	89 %

1) – Attribute "building land-use" does not participate in classification;
2) – Attributes "weather" and "building land-use" do not participate in classification;
3) – Attributes "day of week", "weather", and "building land-use" do not participate in classification.

The classification results of transportation mode choice slightly worsened after the third attribute "day of week" had been excluded from the dataset. This can be due to the day of the week when the data observed is important for transportation mode choice classification despite the information gain results (see Fig. 8), so attribute "day of week" should participate in classification. The total transportation mode choice classification accuracy was 91.8 % for dataset *N3+*.

The ROC curve was used to evaluate the validity of the classification results of the CART and C4.5 decision tree algorithms for transportation mode choice, that is, how the model is able to correctly classify the transportation modes and visualize the results [46].

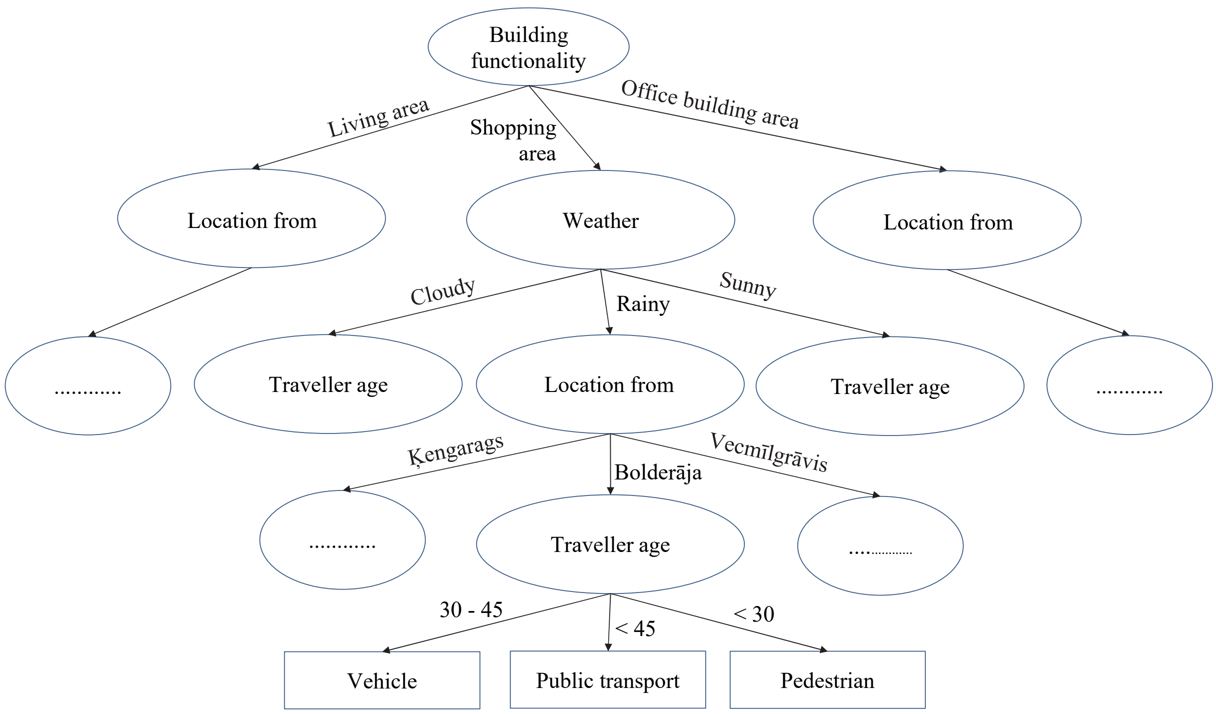


Fig. 9. Example of C4.5 decision tree algorithm results for dataset *N2*.

The application of the ROC curves to the classification results of CART decision tree algorithm (dataset N3+) shows (Fig. 10) that the curve is quite convex and is approaching the top left-hand corner. The AUROC (area under ROC curve) value is close to one, which points to both classification qualities of the model.

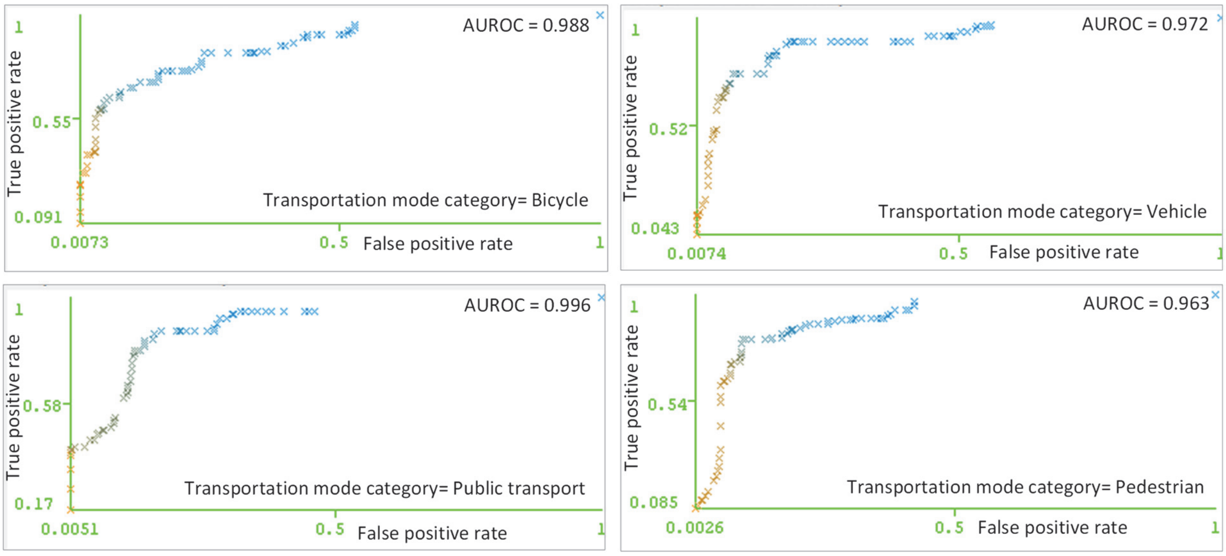


Fig. 10. *ROC* curves for transportation mode categories (dataset *N3+*). Classification performed with *CART* decision tree algorithm.

The application of the ROC curves for classification results of the C4.5 decision tree algorithm shows that also here the curve is quite convex and approaches the top left-hand corner. The AUROC value is close to one for the transportation mode attribute categories “bicycle” and “vehicle”, which indicates a very good model classification quality (see Fig. 11).

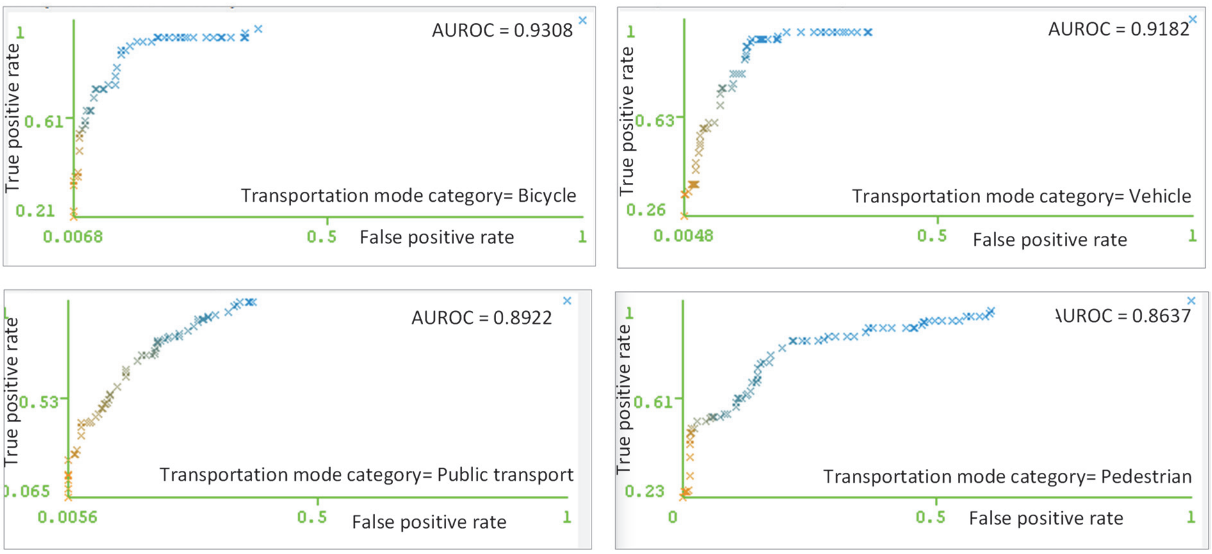


Fig. 11. ROC curves for transportation mode categories (dataset *N3+*). Classification performed with C4.5 decision tree algorithm.

In turn, for transportation mode categories “public transport” and “pedestrian”, the curve is not that convex and close to the top left-hand corner like it does in other traffic mode categories, and the area value under ROC curve is slightly under 0.90, which shows good classification qualities.

To verify the decision tree application efficiency, the transportation mode was classified with traditional transportation mode choice methods for dataset *N3+* (see Table 5).

Table 5

Transportation mode choice results with various methods

Transportation mode choice method	Correctly classified instances
CART decision tree algorithm	**91.8 %**
Discriminant analysis – stepwise method	79.6 %
Discriminant analysis – sequential method	79.6 %
Multinomial logit regression model	88.6 %

As can be seen from Table 5, CART decision tree algorithms provide more accurate transportation mode choice classification results than the multinomial logit regression model and discriminant analysis.

Trip assignment

Dynamic assignment [131] implemented in the integrated transport simulation software has been used to adjust transport routes during the trip and to analyse various transport experiments in transport system over-congested situations.

The conclusions of Section 3 are as follows:

1) *combining various data mining methods for determination of the sample size for transportation mode choice can reduce the number of cases when the model is too well adapted to the data and the noisy data are interpreted as normal;*
2) *information system ITE Trip Generation with Urbemis smart growth method can be used to determine the number of the generated trips for a building object if the surrounding territory of the building has a well-developed transport, public transport, pedestrian and bicycle infrastructure;*
3) *gradient and traversal origin-destination matrices can be used as initial for trip distribution. Experimental results have shown that the initial matrix type does not influence the trip distribution results.*

Section 4 "Transport travel demand simulation model calibration procedure development for transport impact analysis" is dedicated to the transport travel demand simulation model calibration procedure development. The aim of calibration procedures is to adjust the TDSM calibration parameters so that they conform to the observed transport system. It is achieved by estimating the calibration parameters' value with Two-step clustering algorithm taking into account that the Two-step clustering algorithm can fast process a large amount of transport data and is able to work with uninterrupted and categorical variables when the amount of necessary clusters is unknown [8].

Transport travel demand simulation model calibration procedure

The calibration procedure of the transport demand simulation model proposed by the author of the Thesis includes the following stages:

1) data input. The data about traffic flow, geometry of junctions and street network as well as other characteristics of transport system and elements are collected. Time period is defined;
2) initial TDSM development. The development of simulation model is based on gathered transport system elements and characteristics, e.g., width and length of traffic lanes in the model, working patterns of traffic lights, allowed vehicle speed. Origin-destination matrices for different traffic types are defined separately, for passenger and freight vehicle;
3) initial TDSM start with the default calibration parameters and their values. Simulation model was run many times to meet 95 % of confidence interval. To observe the results of modelling, the modelled and observed transport flow intensity is analyzed and root mean square error is defined;
4) if the root mean square error exceeds the defined limit (in the Thesis, 15 % is selected based on the expert's opinion), the evaluation of TDSM parameters' influence on the results is done. A number of experiments with different calibration parameters' values are conducted, and the observed data are compared with modelling results. In the Thesis, the author proposes to use the Two-step clustering for transport travel demand simulation model global parameter value selection;
5) running TDSM with chosen global parameter values. Modelling results are considered acceptable if the root mean square error for the intensity of the modelled and observed flows does not exceed 15 %;
6) planning of TDSM experiments. Necessary changes in the intensity of traffic flow, geometry of street network and transport infrastructure are defined with the aim to determine how the planned traffic changes will affect the behaviour of transport system, identify the potential transport bottleneck, and provide safe and smart transport solutions;
7) the evaluation of the results of TDSM experiment modelling with statistical methods – root mean square error, GEH and Theil's inequality coefficient –, and with transport system effectiveness measures – service level for an analyzed unit – should be equal to D or E (*Approaching unstable or acceptable delay or unstable process, or significant delay*).

In the Thesis, the **proposed** TDSM **calibration procedure** was considered **for transport impact analysis** case study, whose objective was to improve transport accessibility for local drivers and to provide the efficient way for them to get to and from Adazi city centre (hereafter, Adazi example) in case of Main Street peak hours by developing intelligent transportation system measures [151], [152]. The results of the transport impact analysis case study in question were used in the Latvian State Research Program "The next generation of information and communication technologies (NexIT)" within the project "Sensor networks and signal processing applications" (2014–2017).

Transport travel demand simulation model calibration parameters

Global parameters "headway" and "simulation step" [2], 7[6], [96], [107], [134] which determine the whole transport system behaviour were used to check the hypothesis that the development of the transport travel demand simulation model calibration procedure provides the calibration parameter value selection task solution with machine learning methods.

Data collection for simulation model development

During the input data preparation stage, the data about each traffic participant type (vehicles, public transport, pedestrians, and bicyclists), transport flow intensity, street network geometry, public transport lanes, stops and movement intervals as well as traffic lights for controlled junctions for the observed transport situation were collected. The initial observed transport data for Adazi example were compiled using video surveillance data that were collected in the evening peak hour (maximum traffic load hour with highest transport flow intensity in 24 hours) on workdays and at the weekend.

Initial TDSM model development and evaluation

In the initial TDSM development result it was concluded that seven runs for one experiment are necessary to make sure that dispersion between the modelled and the observed transport flow intensities does not exceed 5 % (to meet 95 % confidence interval). Initial TDSM modelled and observed data root mean square error is within 12–28 % depending on the modelled time interval. The significant difference between the modelled and the observed transport flow intensities in Adazi example may also be due to the fact that in the observed situation, vehicles in simulation model were not able to reach the final destination taking into account the blocking situations when one vehicle wants to change the lane and is looking for or waiting for an acceptable gap to make a move, and blocks the lane.

Two-step clustering for global calibration parameter value selection

Two-step clustering is based on maximum likelihood estimation with Akaike (1) Information Criterion or Bayesian (2) Information Criterion [8], [58], [77], [109]. Information criteria are strong tools for model comparison and identification in time series and linear regression. They are widely used in biology, marketing as well as environmental and marine studies [3], [113].

$$AIC = -2\sum_{j=1}^{J}\xi_j + 2m_j \, , \tag{1}$$

$$BIC = -2\sum_{j=1}^{J}\xi_j + m_J \log(N) \, , \tag{2}$$

$$m_J = J\left\{2K^A + \sum_{k=1}^{K^B} L_k - 1\right\} , \tag{3}$$

where AIC – Akaike Information Criterion value;
m_j – number of evaluated parameters, number of free parameters;
ξ_j – likelihood function value dataset;
BIC – Bayesian Information Criterion value;
N – value of training data set;
L_k – number of category for *k* categorical variable;
K^A – total continuous variable number;
K^B – total number of categorical variable.

Two-step clustering results for Adazi example have shown that the optimal number of clusters is three. The centroids and frequencies for considered variables for calibration procedure for Adazi example are given in Table 6.

Table 6

Two-step clustering results

Number of clusters	Centroids				
	Travel time, sec	Observed vehicles, unit per minute	Simulated vehicles, unit per minute	**Simulation step, sec**	% of vehicle O-D by route
1	82.64	11.90	12.41	**0.75**	100.00
2	100.21	3.68	3.50	**0.74**	90.66
3	20.53	19.80	37.16	**0.73**	12.87

It can be seen from Table 6 that in the cluster, the centroid value for simulation step is within the 0.73–0.75 range. The first cluster is the medium flow condition that contains 11–12 vehicles per one minute, and for all these vehicles had only one route from origin to destination (% of vehicle O-D by route – 100 %), and 82.6 % of vehicles had uniform distribution. The second cluster is low flow condition that contains approximately four vehicles per one-minute time interval, mostly 90 % of vehicles had one route from origin to destination, and 67.3 % of vehicles had normal distribution. The third cluster is high flow conditions that contain 37–39 vehicles per one-minute time interval, only 12.8 % of vehicles had one route from origin to destination (% of vehicle O-D by route – 12.8 %), and 91.4 % of vehicles had exponential distribution. The third cluster with high volume of simulated vehicles in comparison with assigned vehicles represents the blocking situation when a too large simulation step is selected.

Hierarchical clustering for the calibration of TDSM total modelled transport system behaviour

In the Thesis, the uniting clustering algorithm based on the researches discussed in [9], [51], [87] was used so as to calibrate the total modelled transport system behaviour. Two distance measures were considered – Manhattan [10] and Mahalanobis distances. To reduce the influence of variables on the clustering solution, ***z*-score** was used **for data normalization** (each variable should have a mean of 0 and a standard deviation of 1). The result of hierarchical clustering is a dendrogram that represents each merge at the similarity between the two merged groups.

Stability of cluster solution

To check cluster solutions for stability, the order of cases was changed. The cases were sorted randomly and for different order of cases, and different cluster solutions were received. There are different methods to define and evaluate the number of clusters [107]. According to the results of the study [84], [85], the Calinski and Harabasz index is the most effective method to determine the optimal number of clusters, followed by the Duda and Hart method (J-index). In the Thesis, the number of clusters (see Table 7) was determined with a distinctive breaking point [87], calculating the variance ratio criterion by the Calinski and Harabasz index and J-index with the SPSS 16.0 and SAS statistical packages.

Table 7

Number of cluster determination results for the Mahalanobis and Manhattan distances

Number of cluster determination methods		**Number of clusters**				
		2	**10**	**11**	**12**	**13**
Distinctive breaking point	Mahalanobis distance	**0.059**	0.24	0.23	0.20	0.18
	Manhattan distance	0.083	0.081	**0.080**	0.61	0.59
Variance ratio criterion	Mahalanobis distance	126.87	272.86	−538.3	452.09	**−554.82**
	Manhattan distance	68.3	−7.68	757.2	**−806.22**	−1.03
J-index	Mahalanobis distance	0.99/ 20.99	**0.53/ 78.88**	0.99/ 1.82	0.97/ 1.59	0.99/ 0.41

Variance ratio criterion [28] is a widely used criterion that computes the ratio of between and within-cluster sums of squares for k clusters. **The J-index** compares the within-cluster sum of squared distance with the sum of within-cluster sum of squared distances and decides whether the cluster should be partitioned into two clusters [84], [85]. The determination of cluster number for hierarchical algorithm showed that the number of clusters varied depending on the chosen method for cluster number determination, and it is definitely impossible to judge about the stability of cluster solution. Hierarchical clustering algorithm can be used in the whole simulated transport system behaviour calibration but requires an in-depth cluster decision stability analysis, which in transport impact analysis context could be inappropriate.

Planning of Experiments

For the Adazi example, the following transport development experiments were planned in order to improve the accessibility of the destination – central part of Adazi city – for local drivers:

1) Experiment *Base* without changes in Street network. "Local" Street has two roundabouts;
2) Experiment *Sc_1* without changes in Street network. At "Local" Street, Variable message sign is placed, which in real-time shows the travel time necessary to cross the Main and Local Streets sections for transit drivers;
3) Experiment *Sc_2*. At "Local" Street, one of roundabouts is changed to a signalized intersection;
4) Experiment *Sc_3*. At "Local" Street, one of roundabouts is changed to a signalized intersection with protected left turns;
5) Experiment *Sc_4*. At "Local" Street, both roundabouts' geometry is changed to signalized intersections with protected left turns;
6) Experiment *Sc_5*. The Variable message sign is placed at "Local" Street entrance. At "Local" Street, both roundabouts' geometry is changed to signalized intersections with protected left turns.

TDSM experiment modelling result estimation

The simulated and observed transport intensities were estimated for normal, uniform and exponential distributions (headway). The evaluation of the transport travel demand simulation model experiments results was done with Theil's uncertainty coefficient, GEH statistics (see Table 8), and root mean square error and simulation models effectiveness measures.

Table 8

GEH statistics results for Adazi example

Experiment	GEH < 5	5 < GEH < 10	GEH > 10
Base	81 %	16 %	3 %
Sc_1	77 %	17 %	6 %
Sc_2	79 %	15 %	6 %
Sc_3	80 %	17 %	3 %
Sc_4	86 %	13 %	<1 %
Sc_5	84 %	14 %	2 %

The root mean square error was within 5–15 % depending on the selected experiments. To pass the GEH test, it is recommended by [132] to have (1) GEH <5 for at least 85 % of links, and (2) GEH within 5–10 for at least 90 % of links. In the considered example, GEH value under 5 provides only the forth transport development experiment. For **Adazi example,** the **effectiveness measures** [39], [132] "travel time for local drivers (to get to city centre)", level of service at intersection and delay time were selected to evaluate the simulation results (see Table 9).

Table 9

Effectiveness measures for experiments

Effectiveness measure	**Experiment**					
	Base	Sc_1	**Sc_2**	Sc_3	**Sc_4**	Sc_5
Travel time for local drivers to get to city centre, sec	214.6	208.1	**205**	208.3	**206**	207.4
Level of service	–	–	**C**	B/C	**B/C**	B

Transport demand simulation model evaluation results with the indexes of effectiveness showed that travel time to the city centre for local drivers varies from 205 to 214 seconds depending on the selected transport development experiment. Transport demand simulation model validation results for Adazi example allows for the conclusion that the developed TDSM model is suitable for future development experiment analysis and evaluation.

The conclusions of Section 4 are as follows:

1) *transport demand simulation model calibration procedure allows solving the task of simulation model calibration adaptation for researched transport system traffic flow influence analysis;*
2) *Two-step clustering allows solving the task of transport demand simulation model global parameter value choice;*
3) *global calibration parameter – transport arriving time gap and use of transport simulation modelling step in the calibration of transport demand model allows estimating the compliance to evaluated transport system;*
4) *the use of statistical methods and two-step clustering algorithm in the calibration of transport demand simulation model enabled processing a large amount of transport data (the results of simulation model), working with continuous and categorical variables and providing the expected modelling results with acceptable specification and accuracy;*
5) *effectiveness index – controlled delay and travel time – allows estimating the transport demand simulation model compliance.*

In Section 5 "Transport travel demand simulation model development for mixed-use building", the transport travel demand simulation model for mixed-use building in the area of Hanzas street was developed. The development of TDSM model is based on the transport demand simulation model development approach worked out in the Thesis, and it is illustrated in Fig. 12.

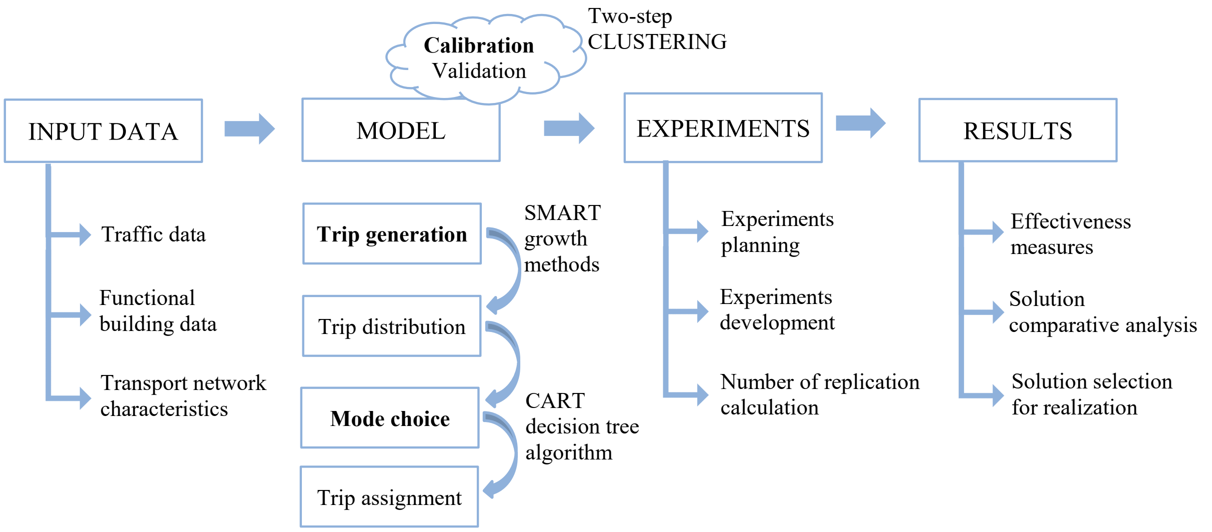


Fig. 12. Transport travel demand simulation model development approach to transport impact analysis.

In **the first stage** of transport demand simulation model **development,** the generated transport trip number was determined using the information system *Trip Generation* together with the smart growth method *Urbemis* [146], [147]. The number of generated transport trips (both incoming and outcoming travels) for mixed-use building will equal 2 258 trips in the evening peak hour (see Table 10), and the final number of generated transport trips has reduced by 40 % in comparison with the initially set according to linear regression equations of the information system *Trip Generation*.

Table 10

Results of the determination of the generated number of transport trips for the evening peak hour for the analysis of traffic flow influence

Functionality	Trip Generation1)	Trip Generation – Automobilization level2)	Trip Generation – Automobilization level – Urbemis3)
	trips/h	trips/h	trips/h
Shopping area	3 005	2 224	1 890
Office space	682	395	312
Total:	3 687	2 619	2 202

1) – The number of generated trips for the maximum evening load hour is determined with the ITE Trip Generation information system;
2) – The number of generated trips for the maximum evening load hour with additional corrections to automobilization level;
3) – The number of generated trips for the maximum evening load hour is determined with Urbemis smart growth methods.

In the **second stage** of TPIM development, the initial gradient origin-destination matrix was used for converting the intensity of traffic flows into an origin-destination matrix [143].

In the **third stage** of TPIM development, the transportation mode choice technique developed by the author of the Thesis was used, where the sample was defined with the minimum volume ellipsoid and the selection of transport mode was done with the CART decision tree algorithm. The target object is a shopping area and an office space of a building. The dataset includes the results of the survey of respondents about their choice of traffic mode [115]. The total size of the dataset is given in Table 11 for sales premises and office space.

Table 11

Total size of the dataset after the functionality of mixed-up building

Functionality	Number of records
Shopping area	2 500
Office space	765

These activities were carried out with the dataset for the definition of the sample, selection of traffic mode, and analysis of the influence of traffic flow:

1) the records with missing data were removed from the dataset; as a result, the dataset decreased by 10 %;
2) the division of dataset attribute according to the values was as follows: cars (33 %), public transport (33 %), and pedestrians (34 %). Records with traffic modes "bicyclists" and "taxi" were removed from the dataset, based on classification analysis about correctly classified record number in the selection of traffic mode, by combining several methods for the definition of the dataset;
3) minimum volume ellipsoid with cost-sensitive evaluation was used to define noisy records in the dataset and to remove them [4], [151], [152]. Pairwise dispersion diagram between the traffic mode and the travel time for the target object – office space – is shown in Fig. 13;
4) choice-based sample and random samples collection was done for the selection of samples for the chosen aim object to provide an even division of traffic mode attribute categories in the dataset.

After the definition of a dataset, the selection of traffic mode was done with the CART decision tree algorithm. Classification accuracy for the selection of transportation mode for target object office space is 90.9 % and for target object shopping area – 92.3 %. Traffic mode division results with regard to target objects are given in Table 12 for the evening peak hour. The travel division results for mixed-use building in the evening peak hour allow concluding that the use of decision tree CART algorithm in the selection of traffic mode was reasonable and provided practicable transportation mode selection results for the transport impact analysis. The difference between the observed and defined with the decision tree

CART algorithm transport trips according to transport modes is 1–3 % for shopping area and 2–5 % for office space.

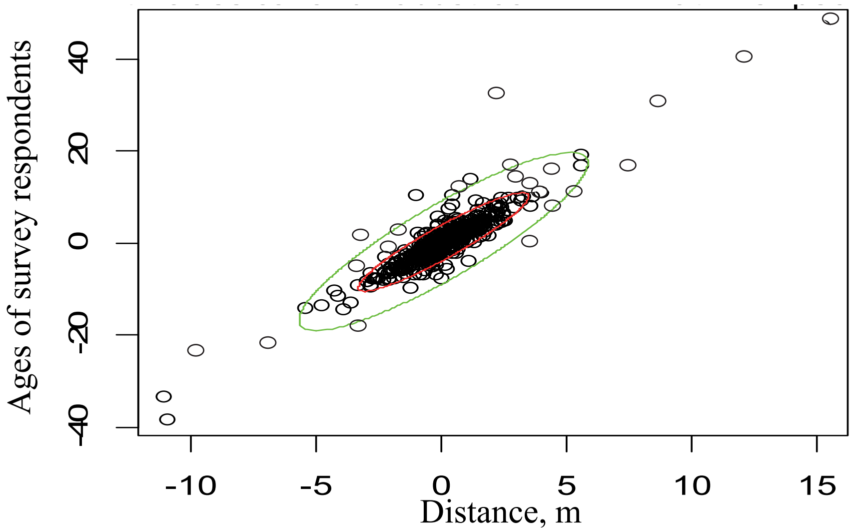


Fig. 13. Minimum volume ellipsoid dispersion diagram for the target object office space.

In the **fourth stage** of TPIM **development,** trip dynamic assignment was carried out.

In the **last stage** of TPIM **development,** the procedure for the definition of calibration methods and model validation developed by the author was used. Root mean square error between the modelled and the observed background transport flow intensity for maximum traffic load hour was 7.9 %, Theil's uncertainty coefficient was 0.031 (the model has good forecasting ability, the intensities of the modelled transport flow are significantly close to the observed ones), and GEH value under 5 % was defined for 89 % of transport street stages.

Table 12

Transportation mode classification results for the evening peak hour

Transportation mode	Shopping area		Office area	
	Calculated with CART[1)]	Collected in the project[2)]	Calculated with CART	Collected in the project
Cars	56 %	57 %	45 %	42 %
Pedestrian	27 %	24 %	9 %	7 %
Public transport	17 %	19 %	46 %	51 %

[1)] – The number of generated trips for the maximum evening load hour is determined with decision tree CART algorithm;

[2)] – The number of generated trips for the maximum evening load hour is determined by the survey [115].

Travel time in the evening maximum load hour for the research territory was calculated for the transport demand simulation model. It allows concluding that the developed transport demand simulation model is successfully validated and can be used to define the necessary traffic changes on the border streets after the building construction.

Planning of the experiments and definition of effectiveness measures

After the development and validation of transport travel demand simulation model, a series of experiments for mixed-use connection to external street network was conducted. The aim of the experiments was to find a solution that would decrease the negative impact of traffic on the transport street network connected to new building on the research territory. The solution foresees that mixed-use building connection for P. Brieža Street will have three lanes with different directions and the junction will be signalled (the necessity of a signalled junction is defined according to LVS 370:2010 "Traffic lights").

Various experiments were carried out for the solution; each of them contained an analysis of different traffic light modes for the junction. A description of the experiments is provided in Fig. 14. The effectiveness measures of six experiments were calculated:

(1) average control delay for cars, trucks, and public transport,
(2) average queue for cars, trucks, and public transport,
(3) average driving speed for cars, trucks, and public transport,
(4) average travel time for cars, trucks, and public transport.

The time of green traffic light signals chosen for the experiment for the street junctions on P. Brieža Street and mixed-use building connection is depicted in Fig. 14. The colour in Fig. 14 reflects common (for cars, trucks, and public transport) average queue. The green colour reflects a short queue for one lane (up to five cars) whereas darker colour means a longer queue.

Fig. 14. Results of the experiments, average transport queue (cars/hours).

Based on the results of the effectiveness indexes analysis, it can be concluded that the second experiment (*Eks_2*) provides less influence of mixed-use building traffic on the transport system in the research territory; therefore, it can be recommended for implementation.

The conclusions of Section 5 are as follows:

1) the developed transport demand simulation model for mixed-up building in the area of Hanzas Street for the evening maximum load hour allowed evaluating the traffic flow intensity connected to mixed-up building and building traffic influence on the transport system on the research territory;
2) the results of the experiments for the developed transport demand simulation model allow concluding that out of six developed experiments the second one provides less average signalled delay and less average queue in the mixed-up building street stage for cars, trucks, and public transport in the research territory.

FINDINGS AND CONCLUSIONS

In the Doctoral Thesis, a transport travel demand simulation model development approach that uses machine learning and simulation methods for transportation mode choice and simulation model calibration parameter adjustment was developed. The developed approach is approved in transport impact analysis for mixed-use building.

Several results were obtained during the development of the Thesis.

1. Transport travel demand models types were studied, and their application in transport impact analysis was analysed.
2. Validation procedures of the transport travel demand simulation model for transport impact analysis were analysed.
3. Analysis of the usability of machine learning and simulation methods in transport demand models was carried out.
4. Machine learning and simulation methods for transport travel demand simulation model development were defined.
5. Transport travel demand simulation model transportation mode choice technique including traveller transportation mode preferences during the trip was developed.
6. Transport travel demand simulation model calibration procedure for calibration parameter value selection with clustering algorithms was elaborated.
7. Transport travel demand simulation model for mixed-use building was constructed.

Based on research results, conclusions were drawn.

1. A comparative analysis of transport travel demand model types has shown that the trip-based transport travel demand model for transport impact analysis requires the application of a new technique and procedure that would include traveller transportation mode preferences during the trip, provide practical usable simulation results accuracy, and satisfy the requirements of transport impact analysis.
2. Application of smart growth *Urbemis* method enables the adjustment of transport generated trips taking into account public transport, pedestrians, and bicycles in the research area.
3. Initial type of origin-destination matrix (traffic intensity for the allocation of the road network) does not influence the trip distribution results.
4. The developed transportation mode choice technique for transport impact analysis provides more accurate transportation mode choice classification results than traditional methods.
5. The developed transport travel demand simulation model calibration procedure allows setting the global calibration parameter values for transport impact analysis.
6. Application of two-step clustering algorithm in simulation model calibration allows processing large simulation model transport data volumes and provides the expected simulation results with appropriate accuracy.
7. Hierarchical clustering algorithm can be used in the whole simulated transport system behaviour calibration; it, however, requires an in-depth cluster decision stability analysis, which in transport impact analysis context could be inappropriate.
8. Transport effectiveness measures – delay and travel time − allow checking the transport travel demand simulation model adequacy.
9. Experimental simulation results obtained from the developed transport travel demand model for the territory between Eksporta, Skanstes, and Hanzas Street in Riga city confirm the efficiency of the developed approach.

Having summarized the research results, the prospects of the developed approach can be marked as follows:

1) an in-depth analysis of the possibility of using machine learning methods in the trip assignment of the trip-based transport travel simulation model development;
2) application of the transport travel demand simulation model development approach in tour-based and activity-based transport travel demand models.

REFERENCES

[1] Abrahamsson T. Estimation of Origin Destination Matrices Using Traffic Counts – A Literature Survey. International Institute for Applied Systems Analysis. – 1998.

[2] Abtahi S. M., Tamannaei M., Haghshenash H. Analysis and modeling time headway distributions under heavy traffic flow conditions in the urban highways: case of Isfahan // Transport. – 2011. – Vol. 26(4). – pp. 375–382.

[3] Acquah H. G. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship // Development and Agricultural Economics. – 2010. – Vol. 2(1). – pp. 001–006.

[4] Aelst S. V., Rousseeuw P. Minimum volume ellipsoid // WIREs Comp Stat. – 2009. – Vol. 1 (1). – pp. 71–82.

[5] Ahipaşaoğlu S. J. Fast algorithms for the minimum volume estimator // Journal of Global Optimization. – 2015. – Vol. 62(2). – pp. 351–370.

[6] Al-Deek H. M. Which Method is Better for Developing Freight Planning Models at Seaports – Neural Networks or Multiple Regression // Transportation Research Board. – 2001. – Vol. 1763. – pp. 90–97.

[7] Algers S., Daly A., Widlert S. Modelling travel behaviour to support policy making in Stockholm. P. Stopher and M. Lee-Gosselin. – 1997.

[8] Altas D., Kubas A., Sezen J. Analysis of environmental sensitivity in Thrace region through two-step cluster // Trakia Journal of Science. – 2013. – Vol. 11. – pp. 318–329.

[9] Anable J. 'Complacent Car Addicts' or 'Aspiring Environmentalists'? Identifying travel behavior segments using attitude theory // Transport Policy. – 2005. – Vol. 12. – pp. 65–78.

[10] ArcGIS Modeling spatial relationships (2013). ArcGIS / Internets. – http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000005000000 – Source viewed 29.05.2013.

[11] Arentze T., Timmermans H. ALBATROSS: A Learning Based Transportation Oriented Simulation System. Den Haag: European Institute of Retailing and Services. – 2000.

[12] Ashok K., Ben-Akiva M. Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Path Flows and Link Flows // Transportation Science. – 2002. – Vol. 36(2). – pp. 184–198.

[13] Balakrishna R., Antoniou C., Ben-Akiva M., Koutsopoulos H. N., Wen Y. Calibration of Microscopic Traffic Simulation Models: Method and Application // Journal of Transportation Research Board. – 2007. – Vol. 1999. – pp. 198–207.

[14] Bekhor S., Ben-Akiva M. E., Ramming, S. Evaluation of choice set generation algorithms for route choice models // Annals of Operations Research. – 2006. – Vol. 144 (1). – pp. 235–247.

[15] Bell M. G. H., Iida Y. Transportation Network Analysis. – New York: John Wiley & Sons, 1997. – 226 pp.

[16] Ben-Akiva M. Travel Demand Modeling, Demand & Economics. Transportation Systems Analysis. MIT OpenCourseWare / Internets. – https://ocw.mit.edu/index.htm – Source viewed 12.12.2016.

[17] Ben-Akiva M., Lu L., Xu Y., Constantinos A. An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models // Transportation Research Part C. – 2015. – Vol. 51. – pp. 149–166.

[18] Ben-Akiva M., McFadden D., Abe M., Böckenholt U., Bolduc D., Gopinath D., Morikawa T., Ramaswamy V., Rao V., Revelt D., Steinberg D. Modelling methods for discrete choice analysis // Marketing Letters. – 1997. – Vol. 8(3). – pp. 273–286.

[19] Bera S., Rao K. V. K. Estimation of origin-destination matrix from traffic counts: the state of the art // European Transport \ Trasporti Europei. – 2011. – Vol. 49. – pp. 3–23.

[20] Bernardin V., Conger M. From Academia to Results from Calibration and Validation of First Hybrid Accessibility-Based Model // Journal of the Transportation Research Board. – 2010. – Vol. 2176. – pp. 50–58.

[21] Bert E., Chung E., Dumont A. Urban dynamic origin-destination matrices estimation // World Congress on Intelligent Transport Systems. – New York: ITS Congress Association, 2008.

[22] Bhat C., Guo J., Srinivasan S., Sivakumar A. Comprehensive econometric microsimulator for daily activity-travel patterns // Journal of the Transportation Research Board. – 2004. – pp. 57–66.

[23] Bierlaire M., Crittin F. An efficient algorithm for real-time estimation and prediction of dynamic OD tables // Operations Research. – 2004. – Vol. 52(1). – pp. 116–127.

[24] Bowman J. L., Bradley M., Castiglione J. Making Advanced Travel Forecasting Models Affordable Through Model Transferability // Washington: Transportation Research Board. – 2013.

[25] Bowman J., Ben-Akiva M. Activity-based Disaggregate Travel Demand Model System with Activity Schedules // Transportation Research A. – 2000. – pp. 1–28.

[26] Brackstone M., Antoniou C., Barcelo J., Celik H. Traffic Simulation: Case for guidelines. European Commission, Joint Research Centre Institute for Energy and Transport. – 2014.

[27] Bradley M., Bowman J., Hadi M., Pendyala R., Bhat C., Waller T. Dynamic, Integrated Model System – Jacksonville-Area Application. Transportation Research Board of the National Academies. Washington. – 2014.

[28] Calinski R. B., Harabasz J. A dendrite method for cluster analysis // Communication in Statistics. – 1974. – Vol. 3(1). – pp. 1–27.

[29] Cascetta E. Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator // Transportation Research Part B: Methodological. – 1984. – Vol. 18(4–5). – pp. 289–299.

[30] Casey H. J. Applications to Traffic Engineering of the Law of Retail Gravitation // Traffic Quarterly. – 1955. – Vol. 9(1). – pp. 23–25.

[31] Castiglione J., Bradley M., Gliebe J. Activity-based travel demand models: a Primer. Transportation Research Board. Washington. – 2015.

[32] Chang J. S., Jung D., Kim J., Kang T. Comparative analysis of trip generation models: results using home-based work trips in the Seoul metropolitan area // The International Journal of Transportation Research. – 2014. – pp. 78–88.

[33] Chatterjee A., Khasnabis L., Slade J. Household stratification models for travel estimation // Journal of Transportation Engineering. – 1977. – Vol. 103(1). – pp. 199–213.

[34] Chen A., Chootinan P., Recker W. W. Examining the Quality of Synthetic Origin–Destination Trip Table Estimated by Path Flow Estimator // Journal of transportation engineering. – 2005. – pp. 506–513.

[35] Ciuffo B., Punzo V., Torrieri V. Comparison of Simulation-Based and Model-Based Calibrations of Traffic-Flow // Journal of the Transportation Research Board. – 2008. – Vol. 2088. – pp. 36–44.

[36] Community guide to Development Impact Analysis. M. M. Edwards / Internets. – https://gis.lic.wisc.edu/wwwlicgf/shapingdane/facilitation/all_resources/impacts/CommDev.pdf – Source viewed 20.09.2012.

[37] Daehyon K. Neural networks for trip generation model // Journal of the Eastern Asia Society for Transportation studies. – 2001. – Vol. 2(4). – pp. 201–207.

[38] Dash R., Dash R. Comparative analysis of k-means and genetic algorithm based data clustering // International Journal of Advanced Computer and Mathematical Sciences. – 2012. – Vol. 3. – pp. 257–265.

[39] Department of Transportation Federal Highway Administration / Internets. – https://www.fhwa.dot.gov – Source viewed 15.07.2015.

[40] Detering S., Schnieder L., Schniede E. Two-Level Validation and Data Acquisition for Microscopic Traffic Simulation Models // International Journal on Advances in Systems and Measurements. – 2010. – Vol. 3. – pp. 48–56.

[41] Dowling R., Skabardonis A., Al V. Traffic Analysis Toolbox Volume III – Guidelines for Applying Traffic Microsimulation Modeling Software. Research, Development, and Technology. Turner-Fairbank Highway Research Center. – 2004.

[42] Duong D. D., Saccomanno F. F., Hellinga B. R. Effects of microscopic traffic platform calibration on errors in safety and traffic metrics // Washington: Transportation Research Board, 2011. – pp. 1–21.

[43] EPA (2013). Mixed-Use Trip Generation Model / Internets. – https://www.epa.gov/smartgrowth/mixed-use-trip-generation-model – Source viewed 30.05.2013.

[44] EPA. COMMUTER Model v2.0. Environmental Protection Agency. – 2005.

[45] Ezell S. Explaining International IT Application Leadership: Intelligent Transportation Systems. The Information Technology & Innovation Foundation. – 2010.

[46] Fawcett T. An introduction to ROC analysis // Pattern Recognition Letters. – 2006. – Vol. 27. – pp. 861–874.

[47] Ferdous N., Bhat C. Comparison of Four-Step Versus Tour-Based Models in Predicting Travel Behavior Before and After Transportation System Changes – Results Interpretation and Recommendations. The University of Texas at Austin Center for Transportation Research. – 2011.

[48] Fosgerau M. PETRA – An Activity-based Approach to Travel Demand Analysis // National Transport Models: Recent Developments. – 2002. – pp. 134–145.

[49] Fratar T. J., Voorhees A. M., Raff M. S. Forecasting distribution of interzonal vehicular trips by successive approximations/with discussion // Highway Research Board. – Washington: Transportation Research board. – 2008. – pp. 53–65.

[50] George Papageorgiou and Athanasios Maimaris (2012). Modelling, Simulation Methods for Intelligent Transportation Systems, Intelligent Transportation Systems / Internets. – http://www.intechopen.com/books/intelligent-transportation-systems/modelling-simulation-methods-for-intelligent-transportation-systems – Source viewed 02.09.2016.

[51] Hagel P., Shaw R. The Influence of Delivery Mode on Consumer Choice of University // European Advances in Consumer Research. – 2008. – Vol. 8. – pp. 531–536.

[52] Hague C. G. The Netherlands National Model, 1990: The National Model System for Traffic and Transport. Ministry of Transport and Public Works. The Netherlands. – 1992.

[53] Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. 3rd edition. – USA: Elsevier Inc., 2012. – 678 pp.

[54] Hellinga B. R. Requirements for the Calibration of Traffic Simulation Models // Canadian Society for Civil Engineering. – 1998. – Vol. IVB. – pp. 211–222.

[55] Hensher D. A., Button K. Handbook of Transport Geography and Spatial Systems. – Kidlington: Pergamon/Elsevier Science, 2004. – pp. 127–146.

[56] Hirsh M., Prashker J., Ben-Akiva M. Dynamic model of weekly activity pattern // Transportation Science. – 1986. – Vol. 20(1). – pp. 24–36.

[57] Hourdakis J., Michalopoulos P. G., Kottommannil J. A Practical procedure for calibrating microscopic traffic simulation models // Journal of the Transportation Research Board. – 2003. – Vol. 1852. – pp. 130–139.

[58] Hu S. (2015). Akaike Information Criterion / Internets. – http://www.ncsu.edu. – Source viewed 05.09.2015.

[59] Huisken G., Coffa A. Neural Networks and Fuzzy Logic to Improve Trip Generation Modelling // Gold Coast, Queensland: Institute of Transport Studies. – 2000.

[60] ITE (2013). Institutes of Transportation Engineers / Internets. – http://www.ite.org – Source viewed 20.05.2013.

[61] ITE. Trip Generation Handbook, 2nd edition. – An ITE Recommended Practice. – 2004. – 162 p.

[62] Yaldi G., Taylor M., Yue W. Examining the possibility of fuzzy set theory application in travel demand modelling // Journal of the Eastern Asia Society for Transportation Studies. – 2010. – Vol. 8. – pp. 579–592.

[63] Jovicic G. Activity based travel demand modelling – a literature study. Danmarks TransportForskning. – 2001.

[64] Jovicic G., Hansen C. O. The Orestad Traffic Passenger Demand Model. Aalborg Universitet. – 2001.

[65] Jung-Beom L., Kaan O. Calibration of a Macroscopic Traffic Simulation Model Using Enhanced Simultaneous Perturbation Stochastic Approximation Methodology // Washington: Transportation Research Board. – 2008. – pp. 1–19.

[66] Khalik A., Amal M. A.-T. Prediction Analysis of Trip Production Using Cross-Classification Technique // Al-Rafidain Engineering. – 2006. – Vol. 14(4). – pp. 51–63.

[67] Kim K. O., Rilett L. R. Simplex-based Calibration of Traffic Microsimulation Models with Intelligent Transportation Systems Data // Journal of Transportation Research Board. – 2003. – Vol. 1855. – pp. 80–89.

[68] Kitamura R. A. Stratification Analysis of Taste Variations in Work Trip Mode Choice // Transportation Research. – 1981. – Vol. 15A(6). – pp. 473–485.

[69] Koonce P. Benefits of Intelligent Transportation Systems Technologies in Urban Areas: A Literature Review. Portland State University. – 2005.

[70] Koppelman F. S., Bhat C. A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. U.S. Department of Transportation. – 2006.

[71] Kulkarni A., McNally M. A microsimulation of daily activity patterns // Washington: Transportation Research Board. – 2001.

[72] Lee D. H., Yang X., Chandrasekar P., Parameter calibration for PARAMICS using Genetic Algorithm // Washington: TRB. – 2001.

[73] Lee G., You S., Ritchie S., Saphores J. D., Sangkapichai J. Environmental Impacts of a Major Freight Corridor: A study of the I-710 in California // Journal of the Transportation Research Board. – 2009. – Vol. 2123. – pp. 119–128.

[74] List G., Konieczny L., Durnford C. A best practice truck flow estimation model for the New // Transportation Research Record. – Washington: Transportation research board, 2002. – pp. 97–103.

[75] Liu H., Yu Q., Ding W., Ni D., Wang H., Shannon S. Feasibility Study For Automatic Calibration of Transportation Simulation Models // Boston: Society for Computer Simulation International, 2011. – pp. 87–94.

[76] Ma J., Dong H., Zhang H. M. Calibration of Micro Simulation with Heuristic Optimization Methods // Journal of Transportation Research Board. – 2007. – Vol. 1999. – pp. 208–217.

[77] Ma J., Kockelman K. M. Crash Modeling Using Clustered Data from Washington State: Prediction of Optimal Speed Limits // IEEE Intelligent Transportation Systems Conference – Toronto: IEEE ITSC, 2006. – pp. 1621–1626.

[78] Maher M. Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach // Transportation Research Part B: Methodological. – 1983. – Vol. 17(6). – pp. 435–447.

[79] Mauro R., Giuffrè O., Granà A., Chiappone S. A Statistical Approach for Calibrating a Microsimulation Model for Freeways // WSEAS Transactions on environment and development. – 2014. – Vol. 10. – pp. 496–508.

[80] Meyer M. D., Miller E. J. Urban Transportation Planning, 2nd edition. – Boston: McGraw-Hill, 2001. – 656 p.

[81] Merkuryev J., Zenina N., Romanovs A. Intelligent Transport Measures as a Component of Cyber-Physical Systems: Case Study for Adazi City // Genova: DIME Universita di Genova, 2015. – pp. 57–65.

[82] Metropolitan Travel Forecasting, Current Practice and Future Direction. Transportation Research Board. – 2007.

[83] Miller E. J., Roorda M. J., Carrasco J. A. A Tour-Based Model of Travel Mode Choice // Lucerne: 10th International Conference on Travel Behaviour Research. – 2003.

[84] Milligan G. W., Cooper M. C. A study of variable standardization // Journal of Classification. – 1988. – pp. 181–204.

[85] Milligan G. W., Cooper M. C. An examination of procedures for determining the number of clusters in a data set // Psychometrica. – 1985. – Vol. 50. – pp. 159–179.

[86] Milne A., Abley S., Douglass M. Comparisons of NZ and UK Trips and Parking Rates. NZ Transport Agency. – 2009.

[87] Mooi E., Sarstedt A. M. A Concise Guide to Market Research. – Springer-Verlag Berlin Heidelberg, 2011. – 344 p.

[88] Morrow-Jones H. A., Kim M. J. Determinants of Residential Location Decisions among the Pre-Elderly in Central Ohio // Journal of Transport and Land Use. – 2009. – Vol. 2(1). – pp. 47–64.

[89] NCTRP. Worksite trip reduction model and manual, Tampa: Florida Department of Transportation. – 2004.

[90] Nelson\Nygaard. Adjusting Site-Level Vehicle Trip Generation Using URBEMIS, San Francisko: Nelson\Nygaard Consulting Associates. – 2005.

[91] Ni D., Leonard II J. D., Guin A., Williams B. M. Systematic Approach for Validating Traffic Simulation Models // Journal of the Transportation Research Board. – 2004. – Vol. 1876. – pp. 20–32.

[92] Olstam J., Tapani A. A Review of Guidelines for Applying Traffic Simulation to Level-of-service Analysis // Procedia Social and Behavioral Sciences. – 2011. – Vol. 16. – pp. 771–780.

[93] Ortuzar J. D., Willumsen L. G. Modelling transport, 4th edition. – John Wiley & Sons Ltd. – 2011. – 586 p.

[94] Otkovič I. I., Tollazzi T., Šraml M. Calibration of microsimulation traffic model using neural network // Expert Systems with Applications. – 2013. – Vol. 40. – pp. 5965–5974.

[95] Ozbay L., Jung-Beom, Kaan. A New Calibration Methodology for Microscopic Traffic Simulation Using Enhanced Simultaneous Perturbation Stochastic Approximation (E-SPSA) Approach // Journal of the Transportation Research Board. – 2009. – Vol. 2124. – pp. 233–240.

[96] Park B., Qi H. Development and Evaluation of Simulation Model Calibration Procedure // Journal of Transportation Research Board. – 2005. – Vol. 1934. – pp. 208–217.

[97] Pavlyuk D., Gromule V. Discrete choice model for a prefered transportation mode // Proceedings of the 10th International Conference "Reliability and Statistics in Transportation and Communication". – Riga: Transport and Telecommunication Institute, 2010. – pp. 143–152.

[98] Paz A., Molano V., Martinez E., Gaviria C., Arteaga C. Calibration of traffic flow models using a memetic algorithm // Transportation Research, Part C. – 2015. – Vol. 55. – pp. 432–443.

[99] Picado R. A Test of Transferability: The SE Florida Activity-Based Travel Demand Model // Columbus: Transportation Research Board. – 2013.

[100] Punzo V., Ciuffo B. How Parameters of Microscopic Traffic Flow Models Relate to Traffic Dynamics in Simulation // Journal of the Transportation Research Board. – 2009. – Vol. 2124. – pp. 249–256.

[101] Quinlan J. R. C4.5: Programs for Machine Learning. – Morgan Kaufmann Publishers Inc. –1993. – 299 p.

[102] Ramos A. L., Ferreira J. V., Barceló J. Modeling & Simulation for Intelligent Transportation Systems // International Journal of Modeling and Optimization. – 2012. – Vol. 2. – pp. 274–279.

[103] Rindt M., McNally G., Craig R. The Activity-based Approach. Department of Civil and Environmental Engineering and Institute of Transportation Studies. Irvine. – 2008.

[104] Ruiter E., Ben-Akiva M. Disaggregate Travel Demand Models for the San Francisco Bay Area // Transportation Research Record. – 1978. – pp. 121–128

[105] Rzeszotko J., Nguyen S. H. Machine Learning for Traffic Prediction // Pułtusk: International Workshop on Concurrency, Specification and Programming. – 2011. – pp. 442–453.

[106] Sabel C. E., Kingham S., Nicholson A., Bartie P. Road Traffic Accident Simulation Modelling – A Kernel Estimation Approach // Dunedin: Spatial Information Research Centre University of Otago. – 2005. – pp. 442–453.

[107] Salvador S., Chan P. Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms // 16th IEEE International Conference on Tools with Artificial Intelligence – Florida: IEEE Computer Society. – 2004. – pp. 576–584.

[108] Samimi A., Razi-Ardakani H., Mohammadian A. A Decision Tree Approach to Analyze Freight Mode Choice Decisions // Proceedings of the 1st European Symposium on Quantitative Methods in Transportation Systems. – Lausanne, Switzerland, 2012.

[109] Şchiopu D. Applying TwoStep Cluster Analysis for Identifying Bank Customers' Profile // Buletinul UniversităŃii Petrol – Gaze din Ploieşti. – 2010. – Vol. LXII. – pp. 66–75.

[110] Schultz G. G., Rilett L. R. Analysis of Distribution and Calibration of Car-Following Sensitivity Parameters in Microscopic Traffic Simulation Models // Journal of Transportation Research Board. – 2004. – Vol. 1876. – pp. 41–51.

[111] Shteinman D. Two Methods to Improve the Quality and Reliability of Calibrating & Validating Simulation Models // Adelaide: Australian Institute of Traffic Planning and Management. – 2014. – pp. 1–19.

[112] Shukla N., Ma J., Denagamage W., Huynh R., Perez P. Tour-based travel mode choice estimation based on data mining and fuzzy techniques // International Symposium for Next Generation Infrastructure – United Kingdom: University College London. – 2015. – pp. 215–220.

[113] Snipes M., Taylor D. C. Model selection and Akaike Information Criteria: An example from wine ratings and prices // Wine Economics and Policy. – 2014. – Vol. 3(1). – pp. 3–9.

[114] Solvers SIA. ITE rekomendāciju pielāgošana Latvijas (Rīgas) apstākļiem: korekcijas koeficientu aprēķins vairākām funkcionalitātēm, Rīga: SIA Solvers. – 2006. (In Latvian.)

[115] Solvers SIA. Transporta satiksmes un gājēju plūsmu izpēte Daugavas kreisā krastā. Rīgas Domes pilsētas Attīstības departaments. Rīga. – 2014. (In Latvian.)

[116] Spiess H. A maximum likelihood model for estimating origin–destination matrices // Transportation Research Part B Methodological. – 1987. – Vol. 21(5). – pp. 395–412.

[117] Spiess H. Conical Volume-Delay Functions // Transportation Science. – 1990. – Vol. 24(2). – pp. 153–158.

[118] Stopher P. R., McDonald K. G. Trip Generation by Cross-Classification: An Alternative Methodology // Transportation Research Record. – 1983. – Vol. 944. – pp. 84–91.

[119] Tamin O. Z., Willumsen L. Transport demand model estimation from traffic counts // Transportation. – 1989. – Vol. 16. – pp. 3–26.

[120] Tangphaisankun A., Nakamura F., Okamura T. Influences of PAratransit as A Feeder of Mass Transit System in Developing Countries Based on Commuter Satisfaction // Journal of the Eastern Asia Society for Transportation Studies. – 2009. – Vol. 8. – pp. 1341–1356.

[121] Tatineni V. C., Demetsky M. J. Supply Chain Models for Freight Transportation Planning. University of Virginia. – 2005.

[122] Tesema T., Abraham A., Grosan C. Rule mining and classification of road traffic accidents using adaptive regression trees // International Journal of Simulation Systems, Science & Technology. – 2005. – Vol. 6. – pp. 80–94.

[123] Tillema F., Zuilekom K., Maarseveen M., Huisken G. Trip generation and trip distribution: Comparison of neural networks and traditional methods. WCTR Society. – 2004.

[124] Traffic Simulation and Data: Validation methods and Applications // T. Toledo, T. Kolechkina, P. Wagner et al. – New York: CRC Press, 2014. – pp. 141–162.

[125] Train K. Discrete Choice Methods with Simulation. 2nd edition. – Cambridge: Cambridge University Press, 2003. – 388 p.

[126] Transportation Impact Analyses for Site Development: An ITE Proposed Recommended Practice. Tulkojums: SIA Solvers un Open World tulkojumu birojs. – 2007.

[127] Treiber M., Kesting A. Microscopic Calibration and Validation of Car-Following Models – A Systematic Approach // ISTTT Procedia. – 2014. – No. 00. – pp. 1–18.

[128] TRICS (2013). Trips Rate Information Computer System / Internets. – http://www.trics.org/ – Source viewed 30.05.2013.

[129] Trips Database TDB Bureau (2013). TDB / Internets. – http://www.tdbonline.org – Source viewed 30.05.2013.

[130] TRMS (2013). Transport Roads and Maritime Services. Road Projects / Internets. – http://www.rta.nsw.gov.au/roadprojects – Source viewed 30.05.2013.

[131] TSS. Microsimulator and Mesosimulator in Aimsun 6. Transport Simulation Systems. – 2009.

[132] UK Highways. Design Manual for Roads and Bridges. UK Highways Agency. – 1992.

[133] Vasconcelos L., Seco Á., Silva A. B. Hybrid calibration of microscopic simulation models // Procedia – Social and Behavioral Sciences. – 2013. – Vol. 00. – pp. 1–11.

[134] Vasconcelos L., Seco Á., Silva A. B. Hybrid calibration of microscopic simulation models // 16th Meeting of the EURO Working Group on Transportation, Social and Behavioral Sciences Procedia – Switzerland: Springer International Publishing. – 2014.

[135] Vuk G. An innovative approach to activity based travel demand modelling // Aalborg: Trafikdage på Aalborg Universitet. – 2011. – pp. 1–9.

[136] Wang S., Wang F., Qiu R. Traffic Impact Analysis using Microscopic Traffic Simulation Model // Journal of Transportation Systems Engineering and Information Technology. – 2009. – Vol. 9. – pp. 81–96.

[137] Wardrop J. G. Some Theoretical Aspects of Road Traffic Research // Institute of Civil Engineers. – 1952. – Vol. 1(3). – pp. 352–362.

[138] Wiering M., Veenen J. V., Vreeken J., Koopman A. Intelligent Traffic Light Control. Institute of Information and Computing Sciences, Utrecht University. Utrecht. – 2004.

[139] Wills M. A flexible gravity-opportunity model for trip distribution // Transportation Research board. – 1986. – Vol. 20B(2). – pp. 89–111.

[140] Willumsen L. Estimation of an O-D Matrix from Traffic Counts – A Review. Leed: Institute of Transport Studies, University of Leeds. – 1978.

[141] Wilson A. The use of the concept of entropy in system modelling // Operational Research Quarterly. – 1970. – Vol. 21(2). – pp. 247–265.

[142] Xie C., Lu J., Parkany E. Work Travel Mode Choice Modeling Using Data Mining: Decision Trees And Neural Networks // Journal of the Transportation Research Board. – 2003. – Vol. 1854. – pp. 50–61.

[143] Zenina N. Origin destination matrix estimation from traffic counts based on Bi level approach // Kaunas: Kaunas University of Technology, 2013. – pp. 42–45.

[144] Zenina N., Borisovs A. An Overview of Artificial Neural Networks Application in Transportation // Proceedings of 14th International Conference on Soft Computing. – Brno: Brno University of Technology, 2008. – pp. 12–16.

[145] Zenina N., Borisovs A. Artificial neural network for analysis of the traffic flow // Proceedings of 16th International Conference on Soft Computing. – Brno: Brno University of Technology, 2010. – pp. 25–32.

[146] Zenina N., Borisovs A. Clustering algorithm for travel distance analysis // Information Technology and Management Science. – 2013. – Vol. 16. – pp. 85–88.

[147] Zenina N., Borisovs A. Regression Analysis for Transport Trip Generation Evaluation // Information Technology and Management Science. – 2013. – Vol. 16 (1). – pp. 89–94.

[148] Zenina N., Borisovs A. Transportation Mode Choice Analysis Based on Classification Methods // Information Technology and Management Science. – 2011. – Vol. 16. – pp. 49–53.

[149] Zenina N., Merkurjevs J. Analysis of Simulation input and output to compare simulation tools // Mechanics Transport Communication. – Sofia: BTY "Todor Kableshkov", 2009. – pp. V-4 – V-10.

[150] Zenina N., Merkurjevs J. The Impact of Accessibility on Transport Infrastructure within Commercial Site // Proceedings of the 14th International Conference "Reliability and Statistics in Transportation and Communication". – Riga: Transport and Telecommunication Institute, 2014. – pp. 76–83.

[151] Zenina N., Romanovs A., Merkurjevs J. Incoming Generated Traffic Flow Estimation based on Transport Access Design and Level of Service // Proceedings of the WSEAS 8th International Conference on Urban Planning and Transportation. – Tenerife: WSEAS Press, 2015. – pp. 195–201.

[152] Zenina N., Romanovs A., Merkurjevs J. Modelling Based Approach for Attracted Transport Readiness Trips Estimation to the Site // Mathematical Models and Methods in Applied Sciences. – 2015. – Vol. 9 (1). – pp. 410–417.

[153] Zenina N., Romanovs A., Merkurjevs J. Transport Simulation Model Calibration with Two-step Cluster Analysis Procedure // Information Technology and Management Science. – 2015. – Vol. 19. – pp. 49–56.

[154] Zhang G., Wang Y., Wei H., Chen Y. Examining Headway Distribution Models with Urban Freeway Loop Event Data // Journal of the Transportation Research Board. – 2007. – Vol. 1999. – pp. 141–149.

[155] Zhang L., Cirillo C., Xiong C., Hetraku P. Feasibility and Benefit of Advanced Four-Step and Activity-Based Travel. University of Maryland. – 2011.

[156] Zuylen H. V., Willumsen L. G. The most likely trip matrix estimated from traffic counts // Transportation research Part B: Methodological. – 1980. – Vol. 14(3). – pp. 281–293.

[157] Иванов Е. Е., Шустров Д., Перешивкин С. Многомерные статистические методы. Кафедра экономической кибернетики и экономико-математических методов. – 2010. (In Russian.)

[158] Орлова И. В., Половников В. А., Габескирия В. Я., Гармаш А. Н., Гусарова О. М., Михайлов В. Н., Пилипенко А. И. Эконометрика. – Москва, 2005. (In Russian.)