

Illinois State University

ISU ReD: Research and eData

Theses and Dissertations

4-9-2020

Reliability Improvement On Feasibility Study For Selection Of Infrastructure Projects Using Data Mining And Machine Learning

Xi Hu

Illinois State University, xhu1238@gmail.com

Follow this and additional works at: <https://ir.library.illinoisstate.edu/etd>



Part of the [Civil Engineering Commons](#), [Computer Sciences Commons](#), and the [Mathematics Commons](#)

Recommended Citation

Hu, Xi, "Reliability Improvement On Feasibility Study For Selection Of Infrastructure Projects Using Data Mining And Machine Learning" (2020). *Theses and Dissertations*. 1249.
<https://ir.library.illinoisstate.edu/etd/1249>

This Thesis is brought to you for free and open access by ISU ReD: Research and eData. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ISU ReD: Research and eData. For more information, please contact ISURed@ilstu.edu.

RELIABILITY IMPROVEMENT ON FEASIBILITY STUDY FOR SELECTION OF INFRASTRUCTURE PROJECTS USING DATA MINING AND MACHINE LEARNING

XI HU

151 Pages

With the progressive development of infrastructure construction, conventional analytical methods such as correlation index, quantifying factors, and peer review are no longer satisfactory in support for decision-making of implementing an infrastructure project in the age of big data. This study proposes using a mathematical model named Fuzzy-Neural Comprehensive Evaluation Model (the FNCEM) to improve the reliability of the feasibility study of infrastructure projects by using data mining and machine learning. Specifically, the data collection on time-series data, including traffic videos (278 Gigabytes) and historical weather data, uses transportation cameras and online searching, respectively. Meanwhile, a questionnaire was posted for collection of the public opinions upon the influencing factors that an infrastructure project may have. Then, this model implements the backpropagation Artificial Neural Network (BP-ANN) algorithm to simulate traffic flows and generate outputs as partial quantitative references for evaluation. The traffic simulation outputs used as partial inputs to the Analytic Hierarchy Process (AHP) based Fuzzy logic module of the system for the determination of the minimum traffic flows that a construction scheme in corresponding feasibility study should meet. This study bases on a real scenario of constructing a railway-crossing facility in a college town. The research results indicated that BP-ANN was well applied to simulate 15-

minute small-scale pedestrian and vehicle flow with minimum overall logarithmic mean squared errors (Log-MSE) of 3.80 and 5.09, respectively. Also, AHP-based Fuzzy evaluation significantly decreased the evaluation subjectivity of selecting construction schemes by 62.5%. It concluded that the FNCEM model has strong potentials of enriching the methodology of conducting a feasibility study of the infrastructure project.

KEYWORDS: Artificial Neural Networks, traffic flow simulation, Fuzzy Logic, Analytic Hierarchy Processing, feasibility study, infrastructure project

RELIABILITY IMPROVEMENT ON FEASIBILITY STUDY FOR SELECTION OF
INFRASTRUCTURE PROJECTS USING DATA MINING
AND MACHINE LEARNING

XI HU

A Thesis Submitted in Partial
Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Technology

ILLINOIS STATE UNIVERSITY

2020

RELIABILITY IMPROVEMENT ON FEASIBILITY STUDY FOR SELECTION OF
INFRASTRUCTURE PROJECTS USING DATA MINING
AND MACHINE LEARNING

XI HU

COMMITTEE MEMBERS:

Haiyan Xie, Chair

Klaus Schmidt

Pranshoo Solanki

ACKNOWLEDGMENTS

I would like to sincerely thank my thesis committee members Dr. Sally Xie, Dr. Klaus Schmidt, and Dr. Pranshoo Solanki, for their professional guidance and precious advice on this research.

Dr. Xie was the first professor who brought me into conducting research projects during my stay at Illinois State University (ISU). Her smiling and patience on students attracted me a lot when I started taking her class in Fall 2018. Fortunately, since then, I have been happily working with her on a research project funded by the Office of Sustainability at ISU, and this thesis independent study project. Honestly, her professionalism and wisdom significantly inspired and motivated me.

I would also much appreciate Dr. Schmidt for his advice on my curriculum arrangement. Beyond that, it's also worth pointing out that he gave me many valuable recommendations on my thesis when I was doing the proposal presentation for this work.

I would profoundly thank Dr. Solanki for his guidance on conducting his research projects about construction materials, and the opportunities he provided, such as letting me teach his students how to do laboratory activities sometimes. Besides, I also want to express my appreciation for his patience in teaching me how to use some lab equipment for research projects.

Beyond above, there are also sincere thanks to my friends, Mr. Danny Zhang, Mr. Junyu Wu, and Miss Huijie Zhang, Mr. Shuang Tang, and Miss Annan Lin, who helped me analyze the video data for this study.

Of course, I would also gratefully thank my parents, my fiancée, who always give me financial and emotional supports. I love you all.

X. H.

CONTENTS

	Page
ACKNOWLEDGMENTS	i
TABLES	v
FIGURES	vii
CHAPTER I: INTRODUCTION	1
CHAPTER II: LITERATURE REVIEW	4
Analysis Methods for Feasibility Studies	4
Studies of Data Mining and Machine Learning	5
Studies of Artificial Neural Networks	7
Studies of Fuzzy Comprehensive Evaluation Model	9
Gaps in Knowledge	11
CHAPTER III: METHODOLOGY	13
Scope of Work	13
Technical Framework	14
Model Design	16
Back-Propagation (BP) Neural Network Algorithm and Modeling	17
Fuzzy Logic Algorithm and Analytic Hierarchy Process	19
Relevant Software Used in This Study	28
CHAPTER IV: DATA COLLECTION, VERIFICATION, AND ANALYSIS	29

Data Collection	29
Data Verification	30
Verification of Traffic and Meteorological Data	30
Verification of Questionnaire Survey Data	34
Analytics of Traffic and Weather Data	35
A Sampling of Traffic Video Datasets	35
Analytics of Statistical Characteristics of Traffic and Weather Data	38
Survey Data Analysis	53
CHAPTER V: MODEL SIMULATION AND VALIDATION	61
Traffic Flow Simulation and Sensitivity Analysis	61
Principles of ANN Traffic Simulation in MATLAB	61
Implementation of ANN Traffic Simulation in MATLAB	67
Validation of ANN Traffic Simulation in MATLAB	85
Sensitivity Analysis of Model Input Variables	89
Summary of Traffic Flow Simulation and Sensitivity Analysis	91
Fuzzy Comprehensive Evaluation of Proposed Construction Schemes	92
Proposed Construction Schemes Associated with this Study	92
Principles of AHP and FEC in MATLAB	100
Implementation of AHP-based FCE in MATLAB	103
Summary of Comprehensive Fuzzy Evaluation of Construction Schemes	120

CHAPTER VI: RESULTS AND CONCLUSIONS	123
REFERENCES	128
APPENDIX A: LAYOUT OF THE PROPOSED MODEL SYSTEM IN THIS STUDY	140
APPENDIX B: CODING SCRIPT FOR 15-MIN NOP/NOV TIME-SERIES BP NEURAL NETWORK SIMULATION	141
APPENDIX C: ONLINE QUESTIONNAIRE ASSOCIATED WITH THIS STUDY	146

TABLES

Table	Page
1. Saaty's Numerical Rating Rule in Pairwise Comparison Process	22
2. Average Random Consistency Index (RI)	23
3. A Summary of Data	29
4. Validation of Traffic Video and Meteorological Data	30
5. Determining Sample Size from a Given Population	36
6. Video Packages	37
7. Descriptive Statistics of Variables (15 minutes)	39
8. Descriptive Statistics of Variables (30 minutes)	40
9. Descriptive Statistics of Variables (60 minutes)	40
10. Determination of Strength of Correlation Relationship	42
11. Correlation Matrix of Variables (15 Min)	43
12. Correlation Matrix of Variables (30 Min)	44
13. Correlation Matrix of Variables (60 Minutes)	44
14. Periodical Train Blocking Events	48
15. Commonly Used Ways of Normalization	63
16. Coding Commands for the Three Normalization in MATLAB	63
17. Determination for Network's Parameters	69
18. Performance Parameters of Model Simulations	84
19. Performance Parameters with Additional Testing for All the Simulations	86
20. Methods of Sensitivity Analysis	89
21. Spearman Correlation Coefficients between Inputs and Outputs of 15-min NOP Simulation	91

22. Spearman Correlation Coefficients between Inputs and Outputs of 15-min NOV	
Simulation	91
23. Influential Factors with Construction Schemes	103
24. Results of Pairwise Judgement According to Saaty's Rule	104
25. Weights of Factors and Corresponding Ranking	105
26. Influential Factors and Corresponding Semantic and Quantitative Assessments to	
Construction Schemes	106
27. Statistical Parameters of Different S-NOP Vectors	108
28. Estimated Boundary Values for S-NOP Vectors	108
29. Statistical Parameters of Different S-NOV Vectors	108
30. Estimated Boundary Values for S-NOV Vectors	109
31. Case Studies Regarding the Average Budget and Construction Duration of Railway-	
crossing Projects in the U.S.	112
32. Estimated Boundary Values of Estimated Project Budgets and Construction Durations	112
33. Quantitative Judgement Rules of Construction Schemes	113
34. Completely Quantitative Judgement of the Proposed Construction Schemes	113

FIGURES

Figure	Page
1. Neural Network Transmission and Structure	8
2. Theoretical Framework	16
3. Example of a 3-Layer Perception of Neural Networks	17
4. Hierarchical Structure	21
5. Triangular Membership Function	25
6. Regular Trapezoidal Membership Function	25
7. Special Cases of Trapezoidal Membership Functions	26
8. Gaussian Membership Function	27
9. Screenshots of Traffic Videos	31
10. Website for Access to Weather Data Including Temperature, Wind Speed, and Humidity	33
11. Screenshot of Website for Data of Precipitation and Air Pressure	33
12. Calculation of the Sum of the Video Numbers of Three Packages with Least Videos	37
13. Random Selection of Three Video Packages	37
14. Statistical Mean of Each Variable	41
15. Statistical Maximum of Each Variable	41
16. R Values of Several Pairs of Variables that Have at Least Moderate Correlation Relationship	45
17. Surface Plot of PWT, VWT, and TBT (15 min)	47
18. Surface Plot of PWT, VWT, and TBT (30 min)	47
19. Surface Plot of PWT, VWT, and TBT (60 min)	48

20. Temporal Distribution of NOP, NOV, and NOT (15, 30, and 60 minutes, respectively)	49
21. Temporal Distribution of Pedestrians (15, 30, and 60 minutes, respectively)	50
22. Temporal Distribution of Vehicles (15, 30, and 60 minutes, respectively)	51
23. Temporal Distribution of Trains (15, 30, and 60 minutes, respectively)	52
24. Residence Location of Participants in Map	55
25. Pie Chart of Residence Location of Survey Participants	55
26. Train Blocking Frequency and Block Time	56
27. Ballot for Four Types of Bridge	57
28. Four Types of Pedestrian Bridge	58
29. One Aspect of Social Impacts the Proposed Pedestrian Overpass Have	59
30. Sustainable and Environmental Impacts of the Pedestrian Bridge	60
31. EIA Process and Project Lifecycle	60
32. Application of ANN in MATLAB R2018b	62
33. Basic Layout of Time-Series Neural Network	64
34. The Layout of the Most Optimal Time-Series Neural Network	69
35. Training Performance of 15-Minute NOP Simulation	70
36. Error Histogram of 15-Minute NOP Simulation	71
37. Regressions of 15-Minute NOP Simulation	72
38. Time-Series Response and Error of 15-Minute NOP Simulation	72
39. Training Performance of 15-Minute NOV Simulation	73
40. Error Histogram of 15-Minute NOV Simulation	73
41. Regressions of 15-Minute NOV Simulation	74

42. Time-Series Response of 15-Minute NOV Simulation	74
43. Training Performance of 30-Minute NOP Simulation	75
44. Error Histogram of 30-Minute NOP Simulation	76
45. Regressions of 30-Minute NOP Simulation	76
46. Time-Series Response of 30-Minute NOP Simulation	77
47. Training Performance of 30-min NOV Simulation	78
48. Error Histogram of 30-min NOV Simulation	78
49. Regressions of 30-Minute NOV Simulation	79
50. Time-Series Response of 30-Minute NOV Simulation	79
51. Training Performance of 60-Minute NOP Simulation	80
52. Error Histogram of 60-Minute NOP Simulation	81
53. Regressions of 60-Minute NOP Simulation	81
54. Time-Series Response of 60-Minute NOP Simulation	82
55. Training Performance of 60-Minute NOV Simulation	82
56. Error Histogram of 60-Minute NOV Simulation	83
57. Regressions of 60-Minute NOV Simulation	83
58. Time-Series Response of 60-Minute NOV Simulation	84
59. Time-Series Response of NOP Simulation for Additional Testing	85
60. Time-Series Response of NOV Simulation for Additional Testing	86
61. Performance Parameters of Pedestrian Flow Simulations	88
62. Performance Parameters of Vehicle Flow Simulations	88
63. Negative Consequences of Periodical Standing Cargo Trains	93
64. Proposed Alternatives of Uptown Normal Project	94

65. Screening Results and Recommendation of the Proposed Schemes for Uptown	
Normal Project	95
66. Scheme 1: Pedestrian Overpass	96
67. Scheme 2: Pedestrian Underpass	97
68. Scheme 3: Pedestrian and Vehicle Underpass	99
69. Principles of Conducting AHP in MATLAB	101
70. Principles of Conducting FEC in MATLAB	102
71. AHP Results Generated in MATLAB	105
72. Distribution of Different S-NOP Vectors	107
73. Probability Distribution of Different S-NOP Vectors	108
74. Distribution of Different S-NOV Vectors	109
75. Probability Distribution of Different S-NOV Vectors	109
76. Quantitative Judgement Transmission for MFP and MPD Based on Traffic	
Simulation Results	110
77. MATLAB Function for Determination of Fuzzy Sets	117
78. MATLAB Code for Comprehensive Evaluation of Proposed Construction Schemes	117

CHAPTER I: INTRODUCTION

Infrastructure, as the fundamental facilities and systems serving a country, city or other areas, is composed of public and private physical improvements and services such as public buildings (municipal buildings, schools, and hospitals), transport infrastructure (roads, railroads, bridges, pipelines, canals, ports, and airports), public services (water supply and treatment, sewage treatment, electrical grid, and dams), open spaces (squares, parks, and beaches), and other long-term, physical assets and facilities. Critical public infrastructure, including public works (roads, bridges, water supply/treatment system, etc.), and facilities like hospitals and schools, fails to work or gets updated appropriately may cause significant reflection on the quality of people's life. For instance, the installation of a new pedestrian facility like overpass or underpass drastically decreases traffic congestion and safety hazards at a designated intersection. Because pedestrian flow purposively separates from hybrid traffic flow, then the safety even travels efficiency of the pedestrian will be improved.

In a real infrastructure project, on the one hand, the selection of a specific construction scheme is evaluated and assessed in the early phase of feasibility study. In this phase, it consists of the analysis of technical feasibility, financial, schedule, and operational feasibility instead of exclusively relying on the subjective notion of anyone or any entity. On the other hand, a scientific and compelling decision should make after a deliberateness with adequate theoretical support. Archetypically, in the feasibility study of the majority of infrastructure construction projects, a multivariate analysis method is widely applied to integrate various factors such as cost, aesthetics, social and economic impact, environmental impact, etc. A comprehensive response would give for an in-depth review. To realize this goal in down-to-earth projects, adequate data, including numeric, textual, image, or graphic ones, are required. Accordingly, adaptive techniques should

be brought into then function as reliable methods theoretically and practically to detect the valuable information hidden in dull data and provide credible support for decision-making.

For gathering information for the feasibility study of infrastructure construction projects as the factors that affect decision-making, multiple tools including questionnaire or survey, personal review, documentation review, case studies, focus group, and observation can work in practice. Previous studies (Ziara et al., 2002) indicated that, as one of the well-developed and widely used methods for making decisions, preliminary feasibility studies (PFS) using analytic hierarchy process (AHP) have commonly played of crucial vital role in decision-making concerning the implementation and prioritization of both public and private investment projects. However, a PFS has also shown several limitations, including a relatively long period of the evaluation process, and the bias and inconsistency of AHP-based model. After that, Yun and Caldas (2009) proposed to apply data mining techniques in large-scale infrastructure projects to identify the decision parameters that have an impact on the results of the feasibility study. In the study, machine learning means comprising artificial neural networks and logistic regression analysis introduced to analyze decision parameters such as funding and investment environment.

Due to the apparent limitations on integrating the information mined from analyzing multiple influential factors within a specific infrastructure project, this study investigates the applicability of the Fuzzy-Neural Comprehensive Evaluation Model (FNECM) comprising AHP-based Fuzzy Comprehensive Evaluation model (FCE) and time-series BP Artificial Neural Networks (ANN) as the analytical means for an exhaustive evaluation of construction schemes in the feasibility study of a railway-crossing facility construction project. A considerable amount of valuable data collected through public surveys, documentation review, case studies, and observation from a feasibility study of a pedestrian facility project and used as model inputs after

cleaning and sorting the data among multiple variables. The hypothetical railway-crossing facility will locate at the intersection of South University Street, Beaufort St, and Amtrak railway in Normal, IL, USA. Precisely, this model consists of two sub-models, which are time-series BP neural network traffic simulation model and AHP-based fuzzy comprehensive evaluation model, respectively. In detail, historical daily weather data and traffic-related data used as the model inputs of traffic simulation models. Then, the corresponding simulation model outputs, including pedestrian and vehicle flow, were used as reliable quantitative references to judge the minimum mobility improvement level of each proposed railway-crossing facility should reach. There is also some other criterion, including a budget, social impact, economic impact, safety improvement level of the facility, etc. which affiliated with all the construction schemes prosed. Then, AHP-based FEC was applied to evaluate them, then give an overall grade for each of the proposed engineering plans. However, it's worth pointing out that all the criteria judgments should be transformed from semantic forms to quantitative using appropriate linear numerical rating rule if qualitative evaluations exit—the overall layout of the FNCCEM model attached in Appendix A.

Overall, this study theoretically contributed to the development of the methodology of the feasibility study in infrastructure projects. Technically, the FNCCEM realized a goal of discrete mining information hidden behind various influential factors that affect the decision-making of implementing infrastructure projects. Besides, theoretically, it enriches the reliability and credibility of conducting feasibility study among infrastructure projects in a human-machine interactive way due to the use of computer programming in this study. Beyond above, this study was also a piece of highly interdisciplinary work due to the use of multiple theories and skills, including data mining, machining learning, computer programming, construction design, fuzzy mathematics, project management, etc.

CHAPTER II: LITERATURE REVIEW

Analysis Methods for Feasibility Studies

Come to the point, the simplest explanation of the feasibility study includes two criteria, the cost required and value to attained, used to judge the feasibility of projects (Young, 1970). Later, a more comprehensive definition of a feasibility study put forward. Evaluating the practicability of a scheme or system is designated to objectively reveal the strengths and weakness of an existing business or proposed venture, chances and threats present in the natural environment, resources required to accomplish, and the prospects for success (Justis & Kreigsmann, 1979; Georgakellos & Marcis, 2009). However, relevant study (Abou-Zeid et al., 2007) demonstrated that there was a significant inconsistency in the procedures of feasibility studies among 91 highway public projects in Egypt, due to incomplete and inaccurate data and the lack of organizational and managerial tools in the proceedings of the general feasibility study. Also, Hyari and Kandil (2009) conducted a study that aimed at validating the feasibility study of a previous highway construction project. The research result revealed that there was a significant estimating gap between the actual estimated costs in the feasibility study. Then, they recommended improving the quality of feasibility study for infrastructure projects using methods, including peer reviews, before-and-after feasibility studies, multistage feasibility studies, and unified scope and methodology to create more reliable support for decision-makers.

Similarly, Frechtling et al. (2010) introduced several means to collect data, such as document and portfolio data, survey data, interview data, case study data, and meta-analysis data for developing an evaluation design for National Center for Research Resources (NCRR). Still, it didn't illuminate how the data were applied practically. One more study (Southland Association, 2012) conducted to assess the practicality of implementing green infrastructure practices on private

commercial properties located on or near the Southland Drive corridor. It brought numerous topographic maps like digital surface models of the study area into its work, which enriched the sources of obtaining data for subsequent analysis of the optimal option of project location and relevant activities with the place.

For data treating and project implementation selection, Dikareva and Voytlovskiy (2016) proposed using an assessment algorithm of project efficiency to analyze the effectiveness and financial feasibility of the underground infrastructure construction projects. Later, Elhalwagy, Ghoneem, and Elhadidi (2017) presented finding the correlation between factors in the feasibility study of using piezoelectric energy harvesting floor in building interior spaces by rearranging them according to their properties as either input or output of simulation formula. Additionally, Zhankaziev et al. (2018) proposed to use simulation modeling programs to determine the quantitative features of the main parameters of traffic flow. The framework of performance indicators and project implementation efficiency evaluated by criterion analysis and cost estimate were the determinants of the optimal option of several project implementation plans. Several methodologies of data treating in feasibility studies of various infrastructure projects have raised. However, each method mentioned above has its limitations because they just specified for a specific field of infrastructure projects. In other words, a more generalized data-driven assessment method in the feasibility study of infrastructure projects needed to provide most of the infrastructure decision-makers more rationally and reliably theoretical foundation.

Studies of Data Mining and Machine Learning

Data mining (Fayyad, 1996) systems, which can be categorized as classification, regression, clustering, or descriptive methods according to the task they solve, combine techniques from a crowd of research areas such as statistics or computer science (database systems and machine

learning). Beyond that, the methods can also characterize by the type of knowledge they generate. For instance, connectionist (artificial neural networks), statistics (Naïve Bayes classifier), or logic (decision trees or classification rules) systems are the common ones widely used in the business or engineering domain. In terms of data mining algorithms, they consist of Clustering, K-means, Self-Organizing Map (SOM), Hierarchical Clustering, Graphic Based Methods, Pattern Discovery, Apriori, FP-Growth, LPMiner, K-Nearest Neighbors, Decision Trees, Artificial Neural Networks, Support Vector Machines, Bayesian Networks, and Fuzzy Logic, respectively (Sarigiannis, 2015). Each of the algorithms has applied in the infrastructure field in different ways. For example, Xu (2018) conducted research that brought the clustering analysis algorithm into analyzing the shortest path of the expressway network in China. The research result indicated that clustering analysis has higher accuracy than the construction hierarchy algorithm and CDZ algorithm. That clustering analysis is a significantly informative method for the planning and review of the highway network. However, feasibility studies of infrastructure projects are comprehensive evaluation-based, and clustering analysis is not capable of being extensively utilized. Among all the data mining algorithms, a previous study (Sarigiannis, 2015) found out that ANN has plenty of advantages in its ability to derive information or meaning from large perplexing datasets. First, ANN doesn't rely on data with normal distribution and can process data containing complex relationships. Second, noisy or fuzzy information can be handled by ANN thanks to fault-tolerant computing. Third, it is capable of generalization, so information that is different from the training data can be interpreted, thus representing a "real-world" solution. Before Sarigiannis (2015), Ji et al. (2011) demonstrated that ANN merely performs well when conducting an overall evaluation but has a limitation on assessing a single criterion. Nevertheless, another algorithm, fuzzy logic, can realize the goal of eliminating the restriction of ANN mentioned above. Based on the pros and

cons of both ANN and FCE, and integration of these two algorithms in the practical application becomes realistic because each of them will play a role in eliminating the defects of the other one.

Studies of Artificial Neural Networks

ANN is a multifunctional system like the human brain. In terms of the human mind, it is a highly perplexing, nonlinear, and parallel computer system. It computes entirely different from the conventional digital computer because of the capacity to organize its structural constituents known as neurons to run specific computations (Haykin, 2001). Larsen (1999) indicated that a neuron or nerve cell is a tiny computer that obtains information through its dendrite tree then calculates continuously. When collective input to the neuron is excessive of a certain threshold, the neuron switches from an inactive to the active state, which means, the neuron is firing. The activation of neurons transmits along the axon to the other neurons in the network. A general artificial neural network, as shown in Figure 1. Currently, because of the ability to reproduce and model nonlinear process, this state-of-art algorithm has been broadly applied in various disciplines including system identification and control (traffic control, trajectory prediction, natural resources management), quantum chemistry, pattern recognition (radar systems, face identification, 3D reconstruction), medical diagnosis, finance (French, 2016), data mining (Buscema et al., 2017), visualization, machine translation, etc. Besides, ANN has found application on the acceleration for reliability analysis of infrastructure subject to natural disasters (Nabian and Meidani, 2017) and to predict foundation settlements.

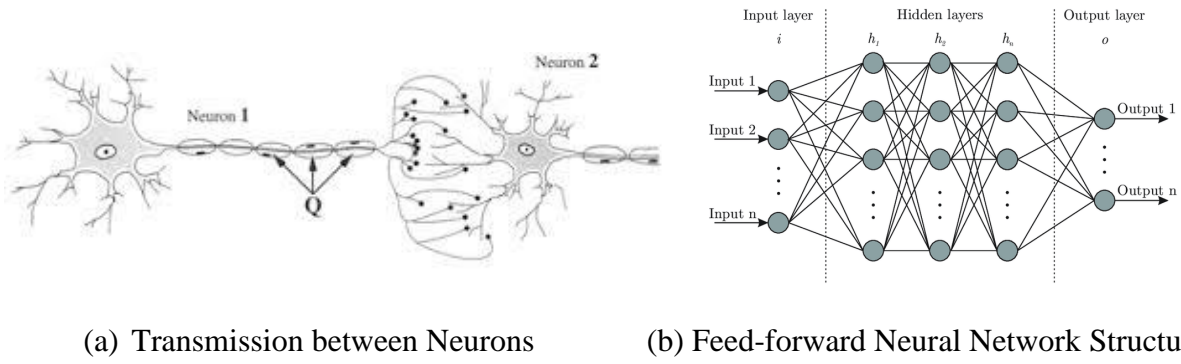


Figure 1 (a)-(b)

Neural Network Transmission and Structure

For the practical use in infrastructure projects, Zhang et al. (2015) proposed to apply ANN for performance evaluation of real estate companies in China from a perspective of stakeholders. They came up with a performance evaluation indicator system to analyze the performance of real estate companies based on the interests of stakeholders. Following the case, the same principle can use to evaluate the performance of contractors in other infrastructure projects.

Similarly, Faye et al. (2016) and Konatowski (2017) successively demonstrated that it is possible to implement artificial intelligence into algorithms for road traffic control and management associated with transportation infrastructure. Subsequently, Brzozowska et al. (2019) and Slimani et al. (2019) brought ANN into the analysis of road traffic management. They deemed that ANN is vital for compiling transport engineering, which contributes to more effective planning, design, and maintenance of transport systems due to ANN has a significantly strong capacity of predicting traffic flow. Lately, an application of ANN and an adaptive network-based fuzzy inference system algorithm presented to investigate the likelihood of green infrastructure enhancement along narrow waterways or on derelict sites (Labib, 2019). At the same time, taking the use of ANN in structural engineering of the bridge as an example, Sitton et al. (2019) proposed an expert system. This system constitutes committees of ANN trained to interrogate data collected

from accelerometers mounted on the railroad bridge then to classify the event as either a passing train or a potentially damaging impact. Concerning the application of ANN in other infrastructure, Ullah et al. (2019) innovatively proposed the use of ANN to build up a smart lighting detection system for smart-city support, which is one of the crucial parameters for building protection. To sum up, ANN plays a vital role as an algorithm to analyze various scientific issues related to nonlinear datasets. But few cases showed that this technique had used to conduct a feasibility study of infrastructure projects for comprehensive assessment at an early date of a specific project.

Studies of Fuzzy Comprehensive Evaluation Model

Classical logic or standard logic (Gamut, 1991) is the intensively studied and most extensively utilized class of logics, but only permits conclusions which are either true or false. In other words, for classical binary logic, the truth of a simple proposition can solely be determined as either 1 (right) or 0 (false). However, a two-valued philosophy has created numerous paradoxes when a statement cannot be both true and false if taking Barber of Seville paradox as an example (Elkind, 2018). Inversely, a fuzzy logic proposition is a statement comprising some concepts without explicitly defined boundaries. Namely, the truth value assigned to this kind of plan can be any value on the interval $[0, 1]$. Gaines (1976) indicated that the fuzzy logic algorithm based on approximate reasoning due to its final objective is to form the theoretical foundation for thinking about the imprecise proposition. Meanwhile, the approximate rationale is analogous to classical logic for reasoning with precise plans.

Around two decades ago, Wang et al. (1999) and Hu (2005) used fuzzy logic in the selection of bridge schemes since it substantially enables to convert qualitative variables such as the difficulty of bridge construction, aesthetics, and social impact to quantitative grey numbers using grey theory. By the end of the last century, Feng and Xu (1999) conceptualized urban

development as a multicriteria decision-making problem. In the study, the layer-by-layer and stage-by-stage fuzzy multicriteria evaluation model was taken advantage of to conduct a comprehensive analysis of urban systems. Analogously, Zhou et al. (2015) proposed an urban regeneration decision-making evaluation system to provide objective and global judgment for urban sprawl and regeneration based on fuzzy logic and entropy weight method, which is an innovative integration of fuzzy logic and other technique in practice. In the transport domain, Zhang et al. (2016) adopted a fuzzy evaluation method with the Analytic Hierarchy Process and Delphi to establish a comprehensive evaluation index system of low carbon road transport. Of which the influential factors included road infrastructure, road traffic, road management, and relevant policies and regulations. Speaking of fuzzy logic applied in project management, Bai et al. (2017) put forward a mathematical model based on fuzzy comprehensive evaluation model and failure mode, effects and criticality analysis for evaluating the sustainability risk level of PPP projects due to majority of studies about PPP projects risk management primarily focus on the impact of financial and revenue risks but ignore the sustainability risks. In the study, a computational experiment associated with a questionnaire survey conducted to verify the feasibility and effectiveness of the model. Similarly, Hendiani et al. (2019) stated that although a large amount of studies has carried out the context of economic and environments in construction projects, less attention always paid to the social aspects. Then, a unique construction social sustainability performance evaluation presented on the strength of fuzzy logic.

In a word, fuzzy logic is an energetically feasible and reliable analytical tool to conduct decision-making issues when a comprehensive evaluation of a specific scheme is needed. Based on the literature review about the fuzzy logic above, it shows that there is a high possibility to combine fuzzy comprehensive evaluation model with other computational algorithms or analytical

methods for a more scientific and persuasive theoretical basis. For instance, Ji et al. (2011) proposed using a fuzzy comprehensive evaluation model and artificial neural networks to evaluate the quality of railway service upon relevant factors including security, punctuality, economy, convenience, and comfortability.

Gaps in Knowledge

Throughout a review of the literature, it can conclude that no work has done to simultaneously bring artificial neural networks and fuzzy logic into the comprehensive evaluation of multiple influential factors and selection of alternatives when various design plans are available. However, the Fuzzy-Neural system has proved working satisfactorily in other fields, such as the evaluation of customers' satisfaction on railway services. Meanwhile, the current methods applied in feasibility studies to integrate all valuable information have specific limitations. For instance, most of them only emphasized on analyzing the one or few out of multiple influential factors including location, structure design, cost, risk, aesthetics, social and economic impact, environmental impact, etc. instead of using a systematic tool to evaluate all the aspects for making decision exhaustively. The current state of practice in feasibility studies of infrastructure projects has no decent resolution in analyzing multiple factors more quantitatively and efficiently. Then, in this study, the primary purpose is to investigate the applicability of data mining and machine learning in updating techniques using in feasibility study for infrastructure projects. Correspondingly, the objectives include but not limited to, the following:

- Explore the application of artificial neural networks (ANN) in small-scale traffic slow simulation.

- Integrate analytic hierarchy processing (AHP) and fuzzy logic for a comprehensive evaluation of the evaluation criterion/influencing factors associated with an infrastructure project.
- Establish the Fuzzy-Neural Comprehensive Evaluation Model (the FNCEM) for optimal selection of construction schemes in the feasibility study of infrastructure projects.
- Enrich the methodologies taken advantage of for conducting a feasibility study of an infrastructure project by tentatively using the rationale of data mining and machine learning.

CHAPTER III: METHODOLOGY

Scope of Work

In this study, as stated, the primary purpose is to remedy the limitations of the current methods for the selection of construction schemes in the feasibility study of infrastructure more quantitatively and intelligently. To be specific, the main points of work in this study show in the following:

1. Collect primary data: collect and mine data from traffic videos. Conduct an online questionnaire to ask the public what elements of the proposed railway-cross facility should have and how they have influenced by periodically standing, and longer. Longer cargo train passing the designated intersection.
2. Collect secondary data: find reliable historical weather data that can describe the meteorological changes at the intersection; conduct relevant case studies for construction design of proposed schemes and determine the evaluation criteria or influential factors.
3. Primarily analyze all data collected to check their statistical characteristics and reveal the potential social issues hidden behind them. Then, propose corresponding solutions for those issues.
4. Build and run a BP neural network traffic simulation model.
5. Constitute and run a comprehensive evaluation of all proposed engineering schemes for those issues.
6. Obtain the overall FNCCEM model.

Technical Framework

In an era of big data, data mining and machine learning have commonly been one of the most popular research interests. Previous studies demonstrated that some conventional methods such as peer review, correlation index of criterion, and quantifying influential factors utilized in the feasibility study of infrastructure projects can't satisfy the continually increasing demand of data analytics for engineering problems. In Figure 2, innovatively, a mathematical computing model named the Fuzzy-Neural Comprehensive Evaluation Model (the FNCEM) is proposed to update the techniques for conducting a feasibility study of construction projects. In this model, fuzzy logic and Back-Propagation (BP) artificial neural networks are the main algorithms implemented in MATLAB to realize the interaction between humans and machines. To activate and validate this model, primary and secondary data collected from a recent feasibility study of pedestrian facility projects. Primary data consisting of survey replies and dynamic videos recording timing movement of both pedestrians and trains in the designated research area, and secondary data retrieved from case studies and external sources are the significant primary inputs for modeling. However, to ensure that appropriate contribution utilized for modeling, all the useful data suffers from verifying authenticity, representativeness, and reliability, respectively. Due to not all raw data, such as dynamic signals of moving subjects in videos can be directly imported to FNESM, data categorization, and preliminary turn to be thoroughly crucial. It's indispensable to clean and filter all the data unless they are ready for being used as input variables for model testing and verification to screen out the data debris from either raw data or secondary data from preliminary analysis. Then, in the end, a sensitivity analysis will be conducted for determining the most influential factors of model output for supporting final decision making of infrastructure project implementation.

Based on the feasible technical route and accessible database and software described ahead, three corresponding hypotheses shown as follows:

- A new machine learning tool, adaptive FNESM, will be successfully developed with applying software both RapidMiner and MATLAB R2018b for mining and comprehensively analyzing multiple influential factors in support for decision-making of engineering problems.
- The dominant factors of final decision-making of implementing an infrastructure project can find out, and the invisible relationship between each element and the final decision will be quantified and displayed visually.
- The observations of this study can be plugged into the existing methodology of conducting a feasibility study of the infrastructure project, systematically explaining the significant applicability of data mining and machine learning in solving practical engineering problems.

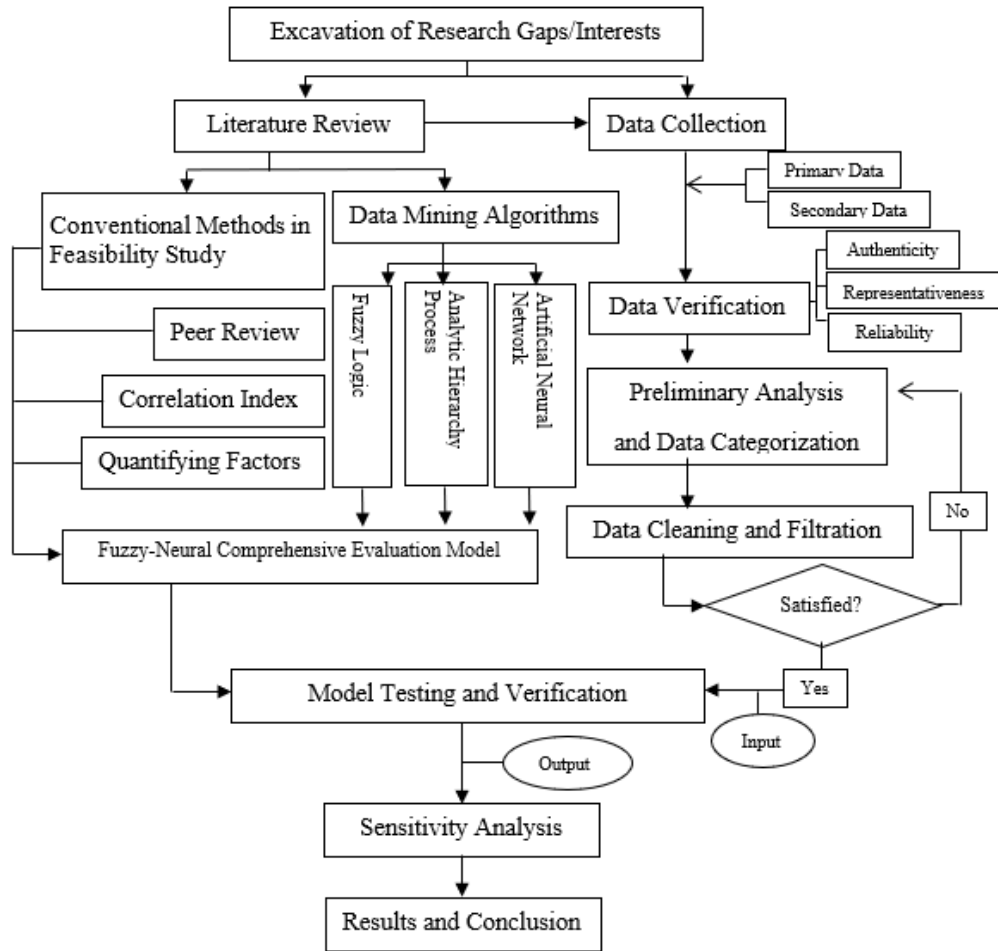


Figure 2

Theoretical Framework

Model Design

Two robust and highly adaptive algorithms, neural networks, and fuzzy logic will introduce in this study for analysis of secondary data generated from the preliminary analysis of row data, including primary and secondary data. Afterward, MATLAB R2018b is supportive of realizing the computerization of these two algorithms.

Back-Propagation (BP) Neural Network Algorithm and Modeling

Back-Propagation neural network algorithm is introduced below due to its strong ability of large-scale computing and unique advantages when dealing with nonlinear high dimensional data (Ding et al., 2010). Detailed procedures are shown in the following:

1. BP Modeling and Relevant Formula

BP algorithm is an iteration-based mathematical computation; the simplified formulas (1)-(3) are shown as follows:

$$y = f((\sum(x_i \cdot W_i^l) + b_o) \quad (1)$$

$$\mathbf{z}^l = \mathbf{W}^l \cdot \mathbf{a}^{l-1} + \mathbf{b}^l \quad (2)$$

$$\mathbf{a}^l = f(\mathbf{z}^l) \quad (3)$$

where y means the predictive model value; $f()$ is a nonlinear function; x_i is the input data point; W_i^l means the weight, and b_o is the offset value or threshold value. Additionally, \mathbf{W}^l means the weight matrix from the $(l-1)^{\text{th}}$ layer to l^{th} layer; \mathbf{a}^{l-1} is the activation value matrix of the $(l-1)^{\text{th}}$ layer; \mathbf{z}^l represents the status value matrix of the l^{th} layer.

In Figure 3, an example of a 3-layer perception of neural networks demonstrates.

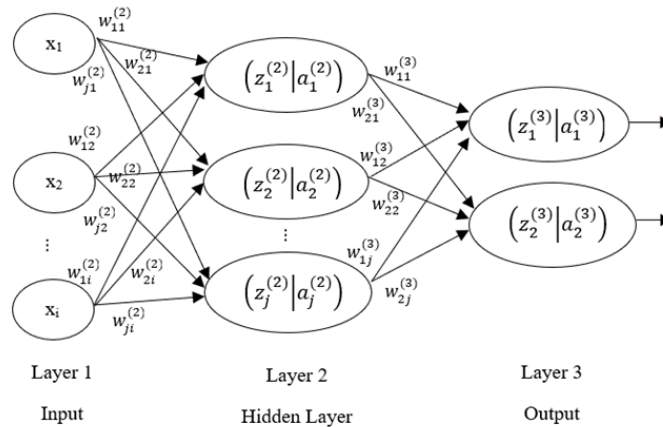


Figure 3

Example of a 3-Layer Perception of Neural Networks

The primary rationale shows in the following steps:

- 1) First of all, continuously calculate the status and activation value for each layer from the first layer to the last one.
- 2) Afterward, calculate the threshold value for each value oppositely from the last layer to the first one.
- 3) And then, update parameters for minimizing the iteration errors and repeat the first two steps mentioned above.

2. Model Solution

It has told that BP neural network comprises at least three essential layers, including input, hidden, and output layer, because there may be multiple sub-layers within the hidden layer. Then, an accurate determination for each segment is given in the following:

1) Design of Input and Output layers

In this study, various indexes containing all vital factors with the preliminary data analysis view as the inputs to this model and the number of nodes (i) is dependent on how many actual parameters would be when conducting the preliminary analysis. The number of output nodes (n) will define as one because of the need for predictive value for each factor.

2) Design of Hidden Layer

Relevant study (Zhong et al., 2017) indicated that a neural network with a hidden layer could approach a nonlinear function with arbitrary precision if there are enough hidden nodes. Hence, with a hidden layer, a 3-layer neural network that has multiple input ports, and only one output port will be chosen as the model framework. An empirical equation used to approximate the number of nodes in the hidden layer presented below:

$$m = \sqrt{i + n} + c \quad (4)$$

where, m means the number of nodes in the hidden layer; i and n respectively represent the number of input nodes and output nodes; c is an integer on an interval $[1, 10]$.

3) Selection of Activation Function

Sigmoid function $S(x)$ will be assigned as the activation function in this study, of which the equation, as shown in the following:

$$S(x) = \frac{e^x}{e^x + 1} \quad (5)$$

4) Implementation of BP Model

Computer programming software MATLAB will use to carry out all the steps described above.

3. Model Validation

In this study, the cross-validation technique will be put into use to validate the predictive results obtained from the ANN model. Because cross-validation is a widely used method to estimate how accurately a predictive model will perform in practice (Zhang et al., 1999).

Fuzzy Logic Algorithm and Analytic Hierarchy Process

A fuzzy comprehensive evaluation, a method based on fuzzy mathematics, absorbs the basic principles of fuzzy logic to quantify the unclearly defined boundaries or qualitative factors. Following the degree of membership, the qualitative evaluation will convert to a quantitative one by avoiding restricted objects or elements. A comprehensive assessment with a fuzzy set will be obtained instead of affirming or denying a result. In total, for a specific evaluation object, there are five steps to establish the modeling of the fuzzy comprehensive evaluation, which shows below:

1. Determination of Factor Set

Assume the factor set as X ,

$$X = (x_1, x_2, x_3, \dots, x_i) \quad (6)$$

where i is the number of evaluation factors.

2. Determination of Evaluation Set

Assume the Evaluation set as Y ,

$$Y = (y_1, y_2, y_3, \dots y_j) \quad (7)$$

where, j is the number of comment levels, and generally lies between 3 and 5.

3. Determination of Weight Vector for Different Factors

Assume A as the weight distribution fuzzy vector,

$$A = (a_1, a_2, a_3, \dots a_m) \quad (8)$$

where, a_m represents the weight of the m^{th} element, and $\sum_{m=1}^n a_m = 1$. Each element in vector A reflects the degree of importance of each corresponding factor.

When running a fuzzy comprehensive evaluation, the weight has a crucial impact on the results. In other words, a significant variation of the pressure may lead to a completely different conclusion. At present, weights for various factors are determined based on experience, which shows that there will be some subjectivity. Generally, the widely used methods for weight determination are constitutive of expert estimation, weighted average, frequency distribution, fuzzy coordination and decision, and analytic hierarchy process (Saaty, 1980; Gunduz et al., 2020). In this study, the analytic hierarchy process (AHP) will be select as the methods for determining the weights for each factor.

4. Principles of AHP

Under Saaty's (1980) definitions of AHP procedures, in total, conventionally, there are six main steps associated with this method when merely applying it for decision-making problems. In this study, since AHP proposes to deal with determining the weights of influential

factors to the final objective of choosing one optimal construction scheme when conducting FEC in MATLAB, in the following, only the first four steps will be detailly discussed.

Step 1: Build up a hierarchical structure. At first, an explicit structure is shown in Figure 4, which contains the overall objective or goal, the relevant criterion that has an impact on the decision-making, and corresponding alternatives should provide for follow-up steps. Note that the elements in either requirements or options layer can be less or more than what shown in this figure. In other words, the specific number of factors in these two layers strictly depends on the practical situation.

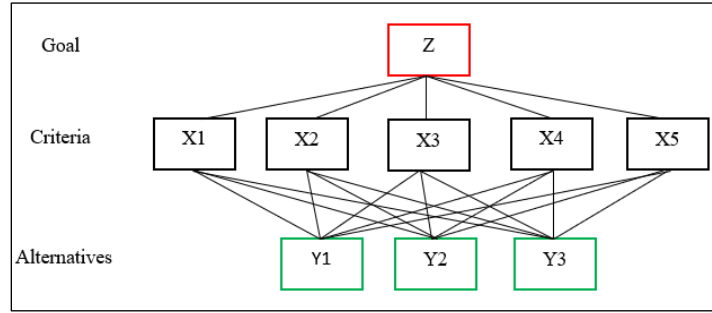


Figure 4

Hierarchical Structure

Step 2: Construct an n-by-n pairwise comparison matrix K . The most significant advantage of AHP is that experts or other sophisticated personnel will judge the importance of paired factors or criteria using Saaty's rule shown in Table 1. Instead of being invited to give a weight vector for all the elements directly. Assume that there are n factors included in a specific evaluation process, the importance of the i^{th} factor over the j^{th} one can note by k_{ij} . Then, an n -by- n pairwise comparison matrix of judgment can obtain according to as follow:

$$K = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ k_{31} & k_{32} & \cdots & k_{3n} \\ \vdots & \vdots & & \vdots \\ k_{n1} & k_{n2} & \cdots & k_{nn} \end{bmatrix} \quad (9)$$

When W_i and W_j respectively represent the weights assigned to factor i and j , three reasonings give below:

(a) $k_{ij} = \frac{W_i}{W_j} > 1$, factor i is more critical than j .

(b) $k_{ij} = 1$, when $i = j$.

(c) $k_{ij} = 1/k_{ji}$, and $k_{ij} \neq 0$.

Table 1

Saaty's Numerical Rating Rule in Pairwise Comparison Process

Numerical Rating	Judgment Preference between Factor i and j
1	Factor i and j have the same importance
3	The factor is slightly important than j
5	The factor is essential than j
7	The factor is powerfully important than j
9	The factor is significantly important than j
2, 4, 6, 8	Intermediate values between two adjacent judgments

Step 3: Calculate the priority vectors w_i and the principal eigenvector value λ_{max} of matrix K . Previous studies (Saaty, 1987) indicated that when $n \leq 3$, an easy way to get an approximation to the priorities is to get the geometric means of the rows normalized. The other method to get this approximation is by normalizing the elements in each column of the judgment matrix then averaging each row. Since the number of factors is generally higher than 3, commonly, the second method is used to get priority vectors. The details of this way show below:

$$w_i = \frac{1}{n} \sum_{j=1}^n \frac{k_{jj}}{\sum_{i=1}^n k_{ij}}, (i, j = 1, 2, \dots, n) \quad (10)$$

Due to principal eigenvector value λ_{max} is a significant index for the subsequent consistency test, it formulates as:

$$\lambda_{max} = \sum_{i=1}^n \frac{(k\bar{w})_i}{nw_i} \quad (11)$$

Step 4: Conduct the consistency test of the judgment matrix. For the sake of ensuring this judgment matrix is reliable, it seems necessary to run the consistency test. Technically, consistency ratio (CR), demonstrate in equation 12, works excellently as a quantitative indicator, which measures how close the judgment matrix is consistent with a purely random form.

$$CR = CI / RI \quad (12)$$

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (13)$$

where CI is calculated from the pairwise comparison matrix using its principal eigenvector value principal eigenvector value λ_{max} . And the size of matrix n. RI represents average random consistency index derived from a capacity of 500, of a randomly generated reciprocal matrix using the scale 1/9, 1/8, ..., 1, ..., 8, 9 to check if the consistency ratio is about to be 0.10 or less. RI corresponding to the size of the judgment matrix, as shown in Table 2.

Table 2

Average Random Consistency Index (RI)

n	1	2	3	4	5	6	7	8	9
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45

The last two of the total six steps will not introduce in this study since the purpose of using AHP only focuses on creating a reasonable and reliable weight vector for the influential factors associated with the decision-making of constructing the railway-crossing facility.

5. Determination of Fuzzy Mapping Matrix R

When only a single factor considered to conduct evaluation then determine its degree membership to assessment set Y, we can conclude that it is a one-way evaluation. After an equal-scale fuzzy subset generated, each factor of a specific evaluation object will be quantified to find out the degree of membership of evaluation object to each different level of the fuzzy subset. Then the fuzzy relation matrix can be obtained as below:

$$\mathbf{R} = \begin{pmatrix} r_{11} & \cdots & r_{12} & \cdots & r_{1n} \\ & & \vdots & & \vdots \\ r_{m1} & \cdots & r_{m2} & \cdots & r_{mn} \end{pmatrix} \quad (14)$$

$$r_i = (r_1, r_2, r_3, \dots, r_n) \quad (15)$$

where r_{ij} represents the degree of membership of one evaluation object from factor set x_j to equal-scale subset y_j ; $\sum_{m=1}^n r_{ij} = 1$.

Concerning the membership function used to calculate the influential factor's degree of membership to each element in the assessment set, it allows us to represent a fuzzy set M graphically. Expressly, assume that fuzzy set M is on the universe of discourse X, then elements x in set M can be defined as $u_M: x \rightarrow [0,1]$, where each element of M maps to a value between 0 and 1 by using corresponding membership function. Widely, triangular, trapezoidal, Gaussian membership functions are used to create fuzzy set M then get a fuzzy matrix integrating all the relevant fuzzy sets.

Below is a relevant theoretical explanation associated with those three membership functions listed above:

Triangular Function: defined by a lower limit a, an upper limit b, and a value x, where $a < m < b$, then the function and the corresponding graph is shown the following respectively.

$$\mu_M(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{m-a}, & a < x \leq m \\ \frac{b-x}{b-m}, & m < x \leq b \\ 0, & x \geq b \end{cases} \quad (16)$$

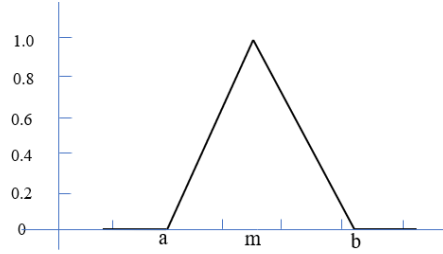


Figure 5

Triangular Membership Function

Trapezoidal Function: defined by a lower limit a, an upper limit d, and a lower support limit b, and a top support limit c, where $a < b < c < d$, then the function and corresponding graph shows in equation (17) and Figure 6, respectively.

$$\mu_M(x) = \begin{cases} 0, & x < a \text{ or } x > d \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \end{cases} \quad (17)$$

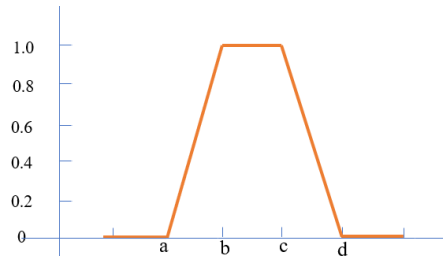


Figure 6

Regular Trapezoidal Membership Function

However, when $a = b = -\infty$, and $c = d = +\infty$, there are two individual cases of trapezoidal function, which are called R-functions and L-functions, and shown in equation (18), and equation (19) respectively.

$$\mu_M(x) = \begin{cases} 0, & x > d \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 1, & x < c \end{cases} \quad (18)$$

$$\mu_M(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases} \quad (19)$$

Additionally, the corresponding graphs, which shows in Figure 7 (a)-(b), are drastically different from the regular trapezoidal shape.

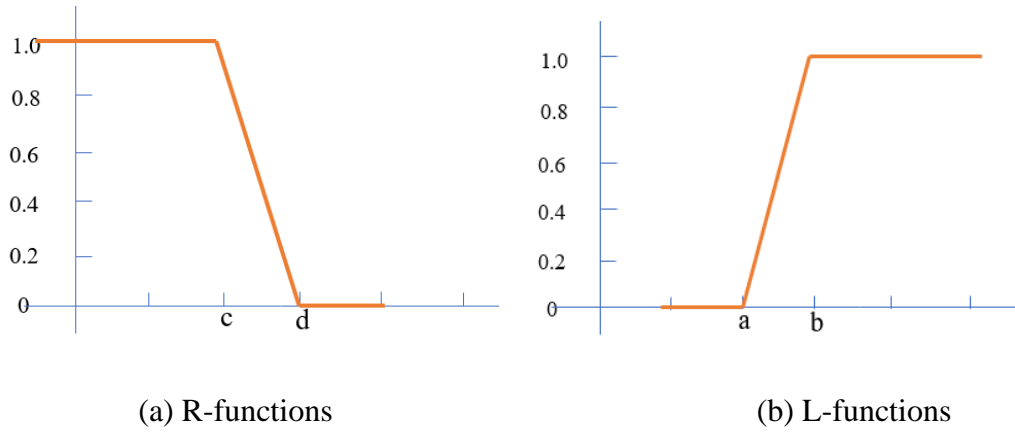


Figure 7 (a)-(b)

Special Cases of Trapezoidal Membership Functions

Gaussian Function: defined by a central value m and a standard deviation $n > 0$. The smaller n is, the narrower the “bell” is. The relevant equation and plotting demonstrated in equation (20) and Figure 8, respectively.

$$\mu_A(x) = e^{-\frac{(x-m)^2}{2k^2}} \quad (20)$$

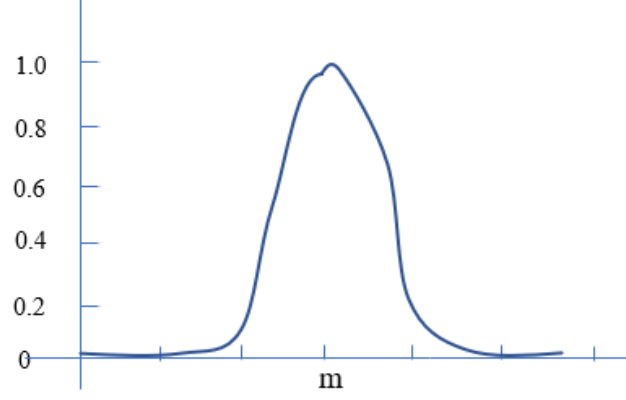


Figure 8

Gaussian Membership Function

6. Implementation of Comprehensive Evaluation

A fuzzy comprehensive evaluation vector **B** can retrieve from a matrix manipulation between fuzzy vector **A** and fuzzy correlation matrix **R**, which shows below:

$$\mathbf{B} = \mathbf{A} \cdot \mathbf{R} = (a_1, a_2, a_3, \dots, a_m) \begin{pmatrix} r_{11} & \cdots & r_{12} & \cdots & r_{1n} \\ \vdots & & \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{m2} & \cdots & r_{mn} \end{pmatrix} = (b_1, b_2, b_3, \dots, b_n) \quad (21)$$

Where b_j is the overall degree of membership of evaluation object to an equal-scale subset y_j

7. Quantitative Analysis of Results

The result of a fuzzy comprehensive evaluation is the degree of membership of evaluation object to its all levels of the fuzzy subset. In general, more information will provide due to it is a fuzzy vector other than a specific value. When several objects evaluated in this model, a further step that is calculating the overall score for each purpose by converting each result vector **B** to a number needs to conduct then sort the scores by size to get the optimal object.

Relevant Software Used in This Study

On technical aspects of applying FCE, Wen (2008) proposed a MATLAB toolbox for grey clustering, which indicated that the soft computing toolbox of grey statistics and the fuzzy comprehension evaluation had developed. Meanwhile, ANN computing can also be conducted in MATLAB, although some other software such as RapidMiner (Ristoski et al., 2015) and Orange (Demšar & Zupan, 2013) are accessible as well. Beyond MATLAB R2018b introduced in this study, Excel 2016 and Minitab 2019 were also used for partial parts of the statistical analysis of data.

CHAPTER IV: DATA COLLECTION, VERIFICATION, AND ANALYSIS

Data Collection

As shown in Table 3, in this study, data can group into two parts, including primary data and secondary data. In detail, one form of primary data collected from a public questionnaire survey about how people responded to the construction of a pedestrian facility over the rail tracks at the intersection of Beaufort Street and South University Street in Normal, IL, USA. Also, more than 1,400 short videos that recorded the traffic movements, including pedestrian crossing behaviors and train passing at the mentioned intersection, were taken. Manual counting and recording may be one of the ways to achieve the image digitalization of video for analysis of pedestrian traffic flow and train traffic flow.

In terms of the secondary data, case studies use to summarize the relevant influential factors like cost, project timing, the satisfaction of aesthetical features, social impact, and investment-revenue analysis, etc. of different completed infrastructure projects for model verification. All the results of the preliminary investigation in this phase will be regarded as secondary data for model simulation and verification subsequently.

Table 3

A Summary of Data

Data Attribute	Data Format	Analysis Method	Resources
Public Survey	Numeral, Polynomial	Statistical Analysis	Qualtrics
Traffic Videos	Dynamic Images	Image Digitalization and Statistical Analysis	Public Works of the Town of Normal
Case Studies	Numeral, Polynomial	Statistical Analysis	Literature Review
Historical Weather Data	Numeral	Statistical Analysis	Online

Data Verification

Verification of Traffic and Meteorological Data

As demonstrated in Table 4, both traffic video data and historical weather data collected in the designated research area have solidly authentic, representative, and reliable based on credible sources of data, including governmental agency and free-to-access commercial websites. These previous research works used these two types of data for traffic simulation. More critically, traffic videos collected for this study have excellent continuity and long recording time from October 19, 2018, to March 22, 2019. At the same time, more than 1400 videos recorded. With regards to the weather data, it has been approved reliable due to its high precision and completeness, and excellent continuity. More details about data verification of these two types of data shown below.

Table 4

Validation of Traffic Video and Meteorological Data

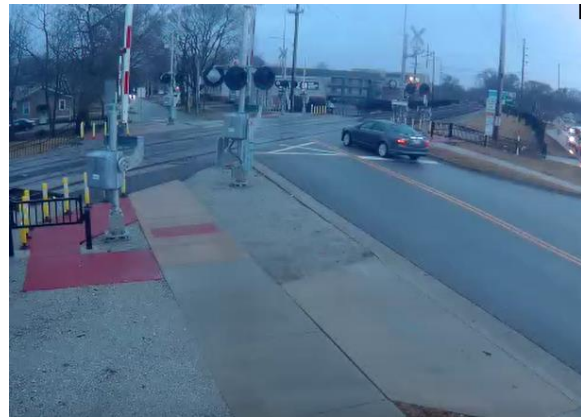
Date Type	Authenticity	Representativeness	Reliability
Traffic Video	Released by governmental agency	Video data are supportive and capable of traffic analysis and simulation. (Thomas S. et al., 2002; Farooq et al., 2019)	Videos have long recording time, excellent continuity, and a high number of videos.
Historical Weather	Retrieved from free-to-access commercial website	The weather has a significant impact on real road conditions and dynamics. (Benjamin et al., 2013; Bie et al., 2017)	High precision; excellent continuity; high completeness

Previous studies (Thomas Schnell et al., 2002; Farooq et al., 2019) revealed that traffic video data are very supportive and capable of being applied to conduct statistical analysis of real traffic situations using 3-D plotting and developing a corresponding appropriate model for traffic simulation. In this study, 1448 traffic videos retrieved from Public Works of the Township of

Normal. Each of them recorded approximately half-an-hour moments among pedestrians, vehicles, and trains. A transportation camera was installed at the intersection of rail tracks, South University Street and West Beaufort Street, IL, the USA by the governmental organization. As shown in Table 4, all 1448 videos collected in chronological order during ten consecutive periods from October 19, 2018, to March 22, 2019. Due to the battery changing, mechanic maintenance, extreme weather like wind chill, and regular pauses for extending camera life expectancy, this camera didn't work all through the whole period mentioned above but recorded at least continuous three-day traffic movements. Additionally, all the videos separately stored in 10 different folders, of which the number of videos in every single folder varied from a minimum of 100 to a maximum of 202.



(a) Night-time Traffic



(b) Daytime Traffic

Figure 9 (a) - (b)

Screenshots of Traffic Videos

To increase the precision of traffic simulation, Benjamin et al. (2013) proposed an artificial neural network forecaster methodology applied to traffic flow condition prediction. In the study, scooter data, timestamp data, and meteorological data utilized as the model inputs. For instance, parameters including temperature, cloud coverage, air pressure, amount of rain, wind

speed, and direction introduced as partial input parameters for the proposed model. Bie et al. (2017) stated that adverse weather indeed has a significant influence on road conditions and traffic dynamics because it lowers the free slow speed, shifts critical density, decreases flow capacity, and contributes to road congestion. In the study, recent historical weather and traffic data used as inputs into real-time traffic prediction and control. Specifically, except for the necessary traffic data, historical weather data were also involved. They consisted of temperature (°F), wind speed (mph), humidity, air pressure (in), and the amount of precipitation (in). They would also be introduced as part of the input variables in this study to obtain more effective results of simulation using a back-forward artificial neural network.

In terms of the hour-by-hour weather data such as temperature, wind speed, and humidity in the Town of Normal collected for this study, they were download from www.timeanddate.com. This commercial website (shown in Figure 10) with more than 1 million daily users established in 1998 for providing free-to-access services comprising of time, worldwide weather and climate, etc. However, due to no access to data about air pressure and precipitation on the above website during those video-recording periods, it's found that the other free website www.wunderground.com (demonstrated in Figure 11) provides these two. Solidly speaking, following the historical data stored in those two websites, all the meteorological data described above are consecutive and accurate.

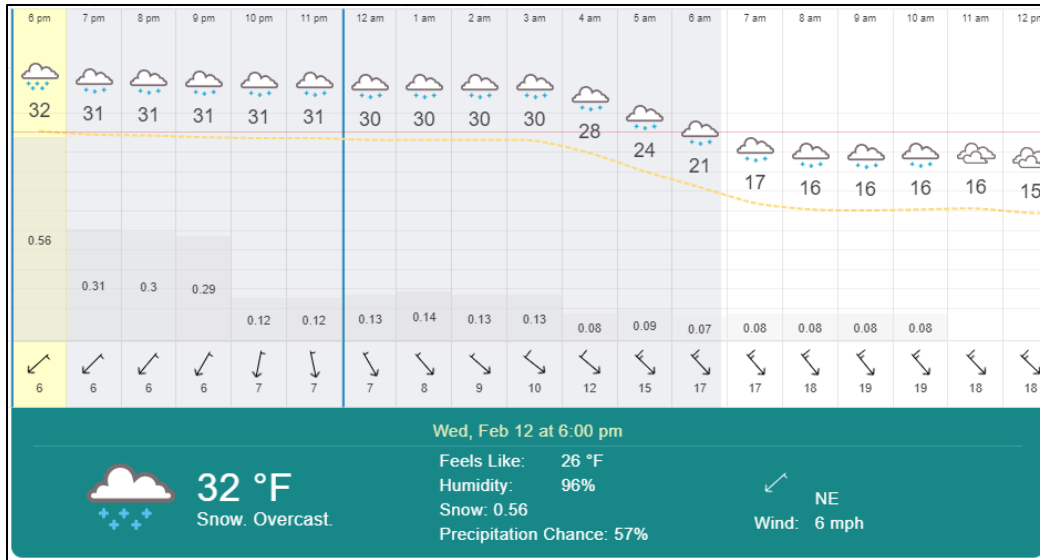


Figure 10

Website for Access to Weather Data Including Temperature, Wind Speed, and Humidity.

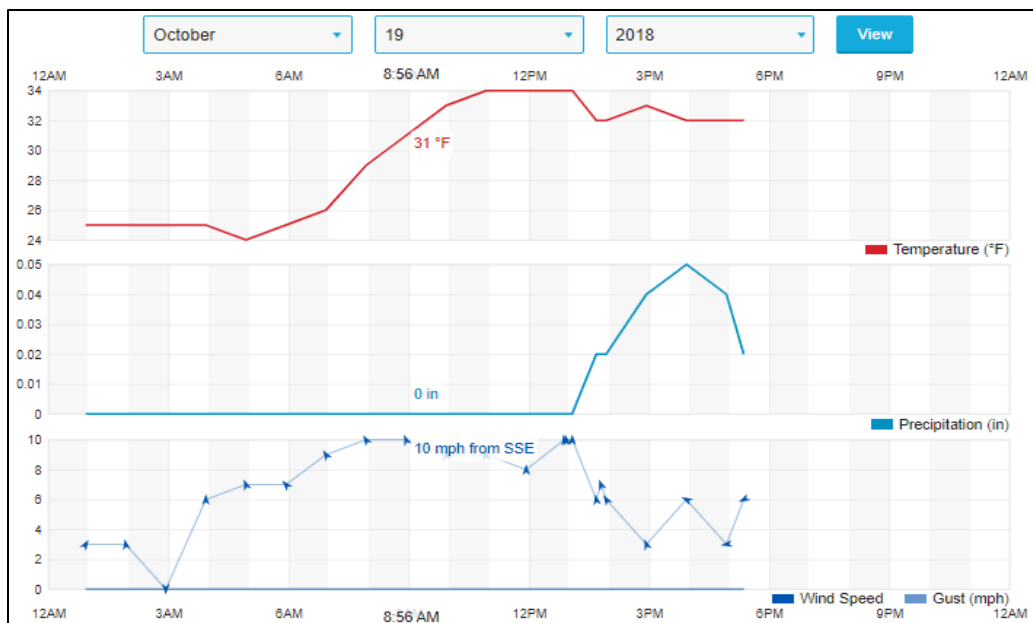


Figure 11

Screenshot of Website for Data of Precipitation and Air Pressure.

Verification of Questionnaire Survey Data

In this study, an online questionnaire which consists of 9 main questions related to a feasibility study of pedestrian bridge project at the selected intersection in this study conducted.

In terms of those questions, they include the information shown in the following:

- Q1: The residence location of survey participators.
- Q2: The feeling (impatient, irritable, or complaining) of pedestrians who have ever been blocked by the trains.
- Q3: How many times pedestrians have ever blocked.
- Q4: How long pedestrians delayed when waiting for the trains to pass by.
- Q5: How necessary the survey participators think a pedestrian overpass should build.
- Q6: What type of bridge they like the best? Such as suspension, truss, arch, or cable-stayed bridge.
- Q7: How the proposed bridge will benefit the people?
- Q8: What the key points of environmental sustainability will be with this bridge.
- Q9: Any other subjective thoughts or comments about this overpass.

Technically, this online survey was dependent on a platform named Qualtrics after the principal investigator got approval from the Institutional Review Board (IRB) of Illinois State University. In other words, this online survey strictly complies with the regulations regarding human subject research following federal rules of the United States. Therefore, the results of this questionnaire have strong authenticity and reliability. In total, 567 responses with different voices received. According to the theory proposed by Krejcie and Morgan (1970), a sample constituting more than 384 elements has the representativeness of a population of one million items, shown in Table 5. Hence, it proved that voices of 567 people in this study adequately

represent the ones of more than 1 million people, indicating significant representativeness of online survey data.

Analytics of Traffic and Weather Data

A Sampling of Traffic Video Datasets

Previous studies (Krejcie and Morgan, 1970) indicated that the sample size could be determined for a specific survey by equation (12) when the population size gives. By this formula, a series of one-to-one correspondence between the estimated sample size and population shows in Table 5.

$$S = X^2 NP(1 - P)/d^2(N - 1) + X^2 P(1 - P) \quad (22)$$

where S is the required sample size; X^2 means the table value of chi-square for 1 degree of freedom at the desired confidence level. N is the population size. P is the population proportion (maximum sample size will provide when it is assumed to be .50); d represents the degree of accuracy expressed as a proportion.

Given that the population size of traffic videos is 1448 in total, shown in Table 6. Approximately, according to the determination of the sample size mentioned above, the sample size should lie between 302 and 306. Since 1448 is significantly close to the average of 1400 and 1500, it shows that the final sample size can also define as an average of 302 and 306, which is 304 exactly. Afterward, more than 304 of 1448 traffic videos are randomly selected and used as the sample to represent the population.

A small program shown in Figure 12 was written in MATLAB R2018b to get enough videos to meet the requirement of sample size in this study, of which the result indicated that at least three video folders should be selected to form the sampling body. In other words, when three

folders that had the least videos sampled, the total number of videos among these three folders was still higher than 304.

Table 5

Determining Sample Size from a Given Population

N	S	N	S	N	S
10	10	220	140	1200	291
15	14	230	144	1300	297
20	19	240	148	1400	302
25	24	250	152	1500	306
30	28	260	155	1600	310
35	32	270	159	1700	313
40	36	280	162	1800	317
45	40	290	165	1900	320
50	44	300	169	2000	322
55	48	320	175	2200	327
60	52	340	181	2400	331
65	56	360	186	2600	335
70	59	380	191	2800	338
75	63	400	196	3000	341
80	66	420	201	3500	346
85	70	440	205	4000	351
90	73	460	210	4500	354
95	76	480	214	5000	357
100	80	500	217	6000	361
110	86	550	226	7000	364
120	92	600	234	8000	367
130	97	650	242	9000	368
140	103	700	248	10000	370
150	108	750	254	15000	375
160	113	800	260	20000	377
170	118	850	265	30000	379
180	123	900	269	40000	380
190	127	950	274	50000	381
200	132	1000	278	75000	382
210	136	1100	285	1000000	384

Table 6*Video Packages*

Date(mm-dd-year)	# of Videos
Uni-Beau 10-19 thru 10-22-18	118
Uni-Beau 10-24 thru 10-26-18	106
Uni-Beau 10-31 thru 11-2-18	100
Uni-Beau 11-7 thru 11-12-18	228
Uni-Beau 12-4 thru 12-6-18	156
Uni-Beau 2-12 thru 2-15-19	156
Uni-Beau 2-20 thru 2-22-19	102
Uni-Beau 2-25 thru 2-28-19	133
Uni-Beau 3-11 thru 3-14-19	147
Uni-Beau 3-18 thru 3-22-19	202
Sum	1448

```

Command Window

>> Nums_videos = [118 106 100 156 156 102 133 147 202];
Min_NumVideos = mink(Nums_videos,3);
>> Min_Sum= sum(Min_NumVideos)

Min_Sum =

    308

fx >>

```

Figure 12*Calculation of the Sum of the Video Numbers of Three Packages with Least Videos*

```

Command Window

>> Nums_videos=[118 106 100 156 156 102 133 147 202];
b = Nums_videos(randperm(length(Nums_videos)));
x=b(1:3)
for i=1:length(b)
    [m,n]=find(Nums_videos==b(i))
end

x =

    147    202    118

```

Figure 13*Random Selection of Three Video Packages*

Then, one more small program demonstrated in Figure 13 generated to determine the random sample comprising of folders with 147, 202, and 118 videos, respectively. Eventually, these 467 videos would use to analyze traffic flow, including pedestrian flow, vehicle flow, and train flow, respectively.

In chronological order, all 467 sampled videos from the population of 1448 played in VCL video player for manual counts of pedestrian flow, vehicle flow, and train flow within time intervals of 15, 30, and 60 minutes, respectively. Besides, traffic congestion indicators such as pedestrian waiting time, vehicle waiting time, and train blocking time, in seconds, were also counted manually during each time lag mentioned above. Ultimately, all the data mined from traffic videos introduced for statistical analysis

Analytics of Statistical Characteristics of Traffic and Weather Data

1. Descriptive Statistics of Model Inputs

Based on the traffic data mined from those sampling videos and weather data collected online, 11 potential variables including pedestrian waiting time (PWT), vehicle waiting time (VWT), train blocking time (TBT), number of pedestrians (NOP), number of vehicles (NOV), number of trains (NOT), and homologous weather parameters consisting of temperature (T), wind speed (WS), humidity (H), atmospheric pressure (AP), amount of precipitation such as rainfall and snow (AOR/S), were defined respectively. Due to the subsequent traffic simulations strictly rely on the data retrieved during those three different time intervals of a given period, after that, three corresponding analytics of descriptive statistics of each variable conducted. In turn, Tables 7, 8, and 9 demonstrate the statistical parameters including mean, standard deviation, minimum, maximum, mode, and the number of patterns of all the 11 variables for 15-minute, 30-minute, and 60-minute intervals. Visually, Figure 14 shows that the means of TBT

and NOV increase when expanding time intervals from 15 to 60 minutes. For instance, TBT increases from an average of 48.39 up to 192.2 seconds. However, means of variables consisting of PWT, VWT, NOP have a weakly positive response to the increment of timespan.

Similarly, the maximum of TBT has the most apparent augmenting response from 900 to 2365 seconds as increasing the time interval. Peak values of variables such as VWT, NOP, and NOV have a slightly increasing tendency when extending timespan. More interestingly, parametric means and maximum of weather elements almost have no changes though time interval enlarged in double.

All in all, under the statistical results discussed above, we can conclude that different time interval indeed has a substantial impact on traffic variables like TBT, NOV, and VWT. Also, when increasing timespan from 15 to 60 minutes in double, neither the mean nor maximum of corresponding traffic variables has no linear responses to the unit increment of time. Therefore, in this study, it's significant to manipulate traffic simulation with data collected and mined from each specific time interval if attempting to get the optimal simulation model.

Table 7

Descriptive Statistics of Variables (15 minutes)

Variable	Mean	St. Dev	Minimum	Maximum	Mode	N for Mode
PWT (S)	2.88	33.06	0	565	0	567
VWT (S)	5.53	37.47	0	568	0	537
TBT (S)	48.39	141.79	0	900	0	444
NOP	3.89	6.83	0	59	0	202
NOV	16.63	14.33	0	66	0	55
NOT	0.19	0.41	0	2	0	470
T(°F)	44.24	8.55	27.0	58.0	44	48
WS (mph)	12.60	7.48	0.0	37.0	5, 12	44
H (%) / 100	0.72	0.23	0.33	3.39	0.89	43
AP (in)	29.20	0.26	28.590	29.48	29.45	36
AOR/S (in)	0.002	0.0143	0.00	0.10	0	561

Table 8*Descriptive Statistics of Variables (30 minutes)*

Variable	Mean	St. Dev	Minimum	Maximum	Mode	N for Mode
PWT (S)	5.71	46.42	0	565	0	283
VWT (S)	10.97	57.19	0	705	0	255
TBT (S)	95.8	245.90	0	1800	0	181
NOP	7.54	12.56	0	98	0	75
NOV	32.62	27.34	0	106	0	18
NOT	0.37	0.58	0	3	0	197
T (°F)	44.16	8.60	27.00	58.8	44	24
WS (mph)	12.64	7.57	0	37.0	5	24
H (%)/100	0.71	0.18	0.33	1.0	0.89	22
AP (in)	29.20	0.26	28.60	29.48	29.45	20
AOR/S (in)	0.003	0.0164	0	0.10	0	281

Table 9*Descriptive Statistics of Variables (60 minutes)*

Variable	Mean	St. Dev	Minimum	Maximum	Mode	N for Mode
PWT (S)	11.46	65.37	0	565	0	138
VWT (S)	22.01	80.17	0	705	0	115
TBT (S)	192.2	380.20	0	2365	0	59
NOP	15.13	22.97	0	129	0	24
NOV	65.47	52.61	0	202	2	6
NOT	0.74	0.83	0	4	0	65
T (°F)	44.24	8.61	27.0	58.0	44, 54	12
WS (mph)	12.70	7.54	0	37.0	5	12
H (%)	0.70	0.18	0.33	1	0.89	11
AP (in)	29.20	0.26	28.59	29.48	29.43, 29.45	9
AOR/S (in)	0.0028	0.0164	0	0.10	0	141

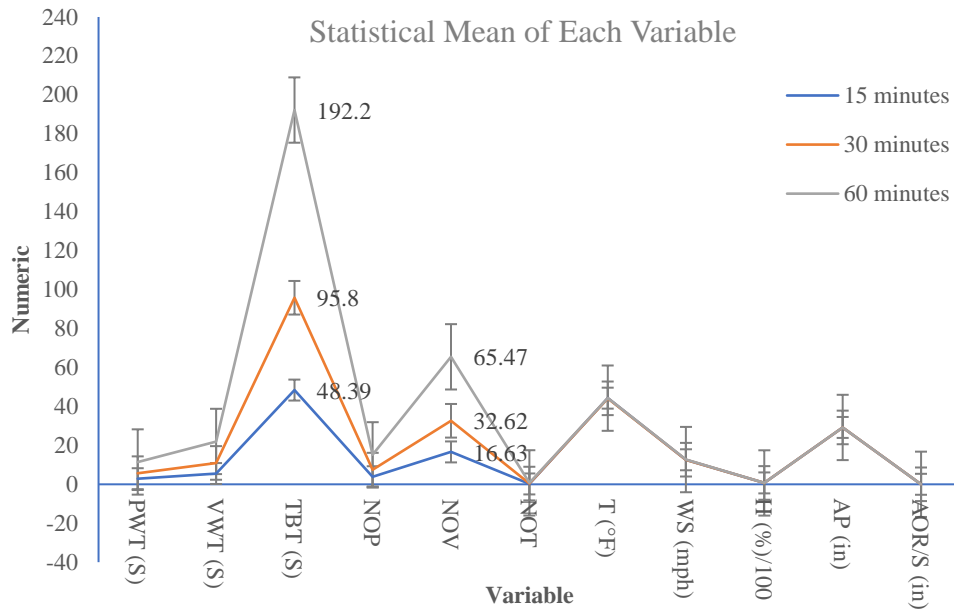


Figure 14

Statistical Mean of Each Variable

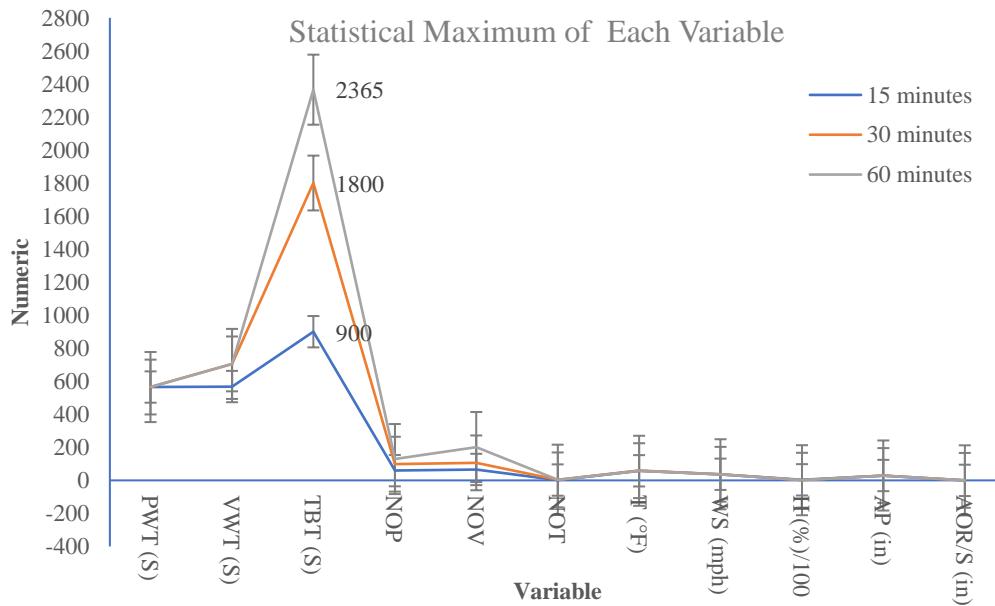


Figure 15

Statistical Maximum of Each Variable

However, to better determine which type of algorithm, linear or nonlinear, for traffic simulations in this study, correlation analysis among variables seems necessary.

2. Correlation Analysis of Model Inputs

Technically, correlation analysis, as a statistical technique commonly used to evaluate the strength of the relationship between two quantitative variables, is strictly connected to the linear regression analysis (Franzese et al., 2019). However, when low strength of linear correlation relationship exists among multiple parameters for modeling, non-linear data mining algorithms can be used. Such as Artificial Neural Networks (Chan et al., 2012), Nonparametric Regression (Zhang et al., 2009), and Support Vector Machine (Fu et al., 2013) may introduce to run non-linear modeling (Lu et al., 2015). In this study, three Pearson correlation analyses conducted to define how strong the correlation relationships among all the 11 variables which comprise of potentially influential model inputs.

As shown in Table 10, it's commonly agreed (Mukaka, 2012; Moore et al., 2013) that the strength of the correlation relationship can categorize into four types based on the absolute value of coefficient R between two variables. $R < 0.3$, $0.3 < R < 0.5$, $0.5 < R < 0.7$, and $R > 0.7$ refer to none or very weak, weak, moderate, and strong strength of correlation relationships, respectively.

Table 10

Determination of Strength of Correlation Relationship

Absolute Value of Coefficient R	Strength of Relationship
$R < 0.3$	None or very weak
$0.3 < R < 0.5$	Weak
$0.5 < R < 0.7$	Moderate
$R > 0.7$	Strong

The results of correlation analyses demonstrated that only a few pairs of variables have a moderate or strong correlation relationship based on the R matrices, as shown in Tables 11, 12, and 13. They are PWT & VWT, NOV & NOP, NOT & TBT, T & WS, T & AP, and WS & AP, have moderate or strong correlation relationship based on the R matrices shown in Table 11, 12, and 13, respectively. Additionally, with an increment of the time interval, these correlation coefficients only vary slightly, as displayed in Figure 16. Then, the rest of the pairs of variables turned to have weak, very weak, or none relationships with each other. Nevertheless, we can't conclude that there are no intercorrelations among the other pairs of variables, although weak, very weak, or none relationships. Most important of all, it's summarized that traffic simulation with linear algorithms is not capable of this study according to the results of the correlation analysis mentioned above. Hence, nonlinear ones such as artificial neural networks or support vector machines have of immediate use.

Table 11

Correlation Matrix of Variables (15 Min)

	PWT	VWT	TBT	NOP	NOV	NOT	T(°F)	WS	H	AP
	(S)	(S)	(S)					(mph)		(in)
VWT (S)	0.789									
TBT (S)	0.236	0.291								
NOP	0.066	0.092	-0.039							
NOV	0.028	0.024	-0.121	0.589						
NOT	0.156	0.275	0.518	-0.005	0.015					
T(°F)	-0.017	0.006	0.000	0.065	0.255	-0.007				
WS (mph)	-0.026	-0.036	-0.010	-0.111	0.009	-0.010	0.660			
H (%) / 100	-0.084	-0.084	0.013	-0.243	-0.389	0.000	-0.119	0.069		
AP (in)	0.074	0.066	-0.011	0.257	0.244	0.021	-0.693	-0.714	-0.361	
AOR/S (in)	-0.013	-0.022	-0.043	-0.083	-0.133	-0.037	0.171	0.151	0.131	-0.278

Table 12*Correlation Matrix of Variables (30 Min)*

	PWT (S)	VWT (S)	TBT (S)	NOP	NOV	NOT	T (°F)	WS (mph)	H	AP (in)
VWT (S)	0.858									
TBT (S)	0.214	0.250								
NOP	0.292	0.288	0.001							
NOV	0.092	0.100	-0.098	0.632						
NOT	0.239	0.301	0.571	0.040	0.054					
T (°F)	-0.013	0.021	0.029	0.060	0.271	0.017				
WS (mph)	-0.019	-0.023	0.013	-0.100	0.041	0.019	0.663			
H (%) / 100	-0.145	-0.128	0.007	-0.387	-0.630	-0.062	-0.242	-0.095		
AP (in)	0.102	0.081	-0.058	0.262	0.205	-0.025	-0.639	-0.659	-0.358	
AOR/S (in)	-0.021	-0.008	-0.050	-0.086	-0.096	0.002	0.228	0.167	0.136	-0.330

Table 13*Correlation Matrix of Variables (60 Minutes)*

	PWT (S)	VWT (S)	TBT (S)	NOP	NOV	NOT	T (°F)	WS (mph)	H	AP (in)
VWT (Min)	0.851									
TBT (Min)	0.153	0.198								
NOP	0.308	0.350	0.021							
NOV	0.162	0.182	-0.050	0.687						
NOT	0.161	0.244	0.685	0.098	0.087					
T (°F)	-0.020	0.028	0.051	0.064	0.278	0.026				
WS (mph)	-0.028	-0.034	0.024	-0.108	0.047	0.024	0.662			
H	-0.206	-0.183	-0.002	-0.422	-0.655	-0.088	-0.247	-0.102		
AP (in)	0.144	0.121	-0.026	0.319	0.278	0.034	-0.699	-0.719	-0.411	
AOR/S (in)	-0.030	-0.047	-0.076	-0.112	-0.175	-0.100	0.202	0.187	0.219	-0.335

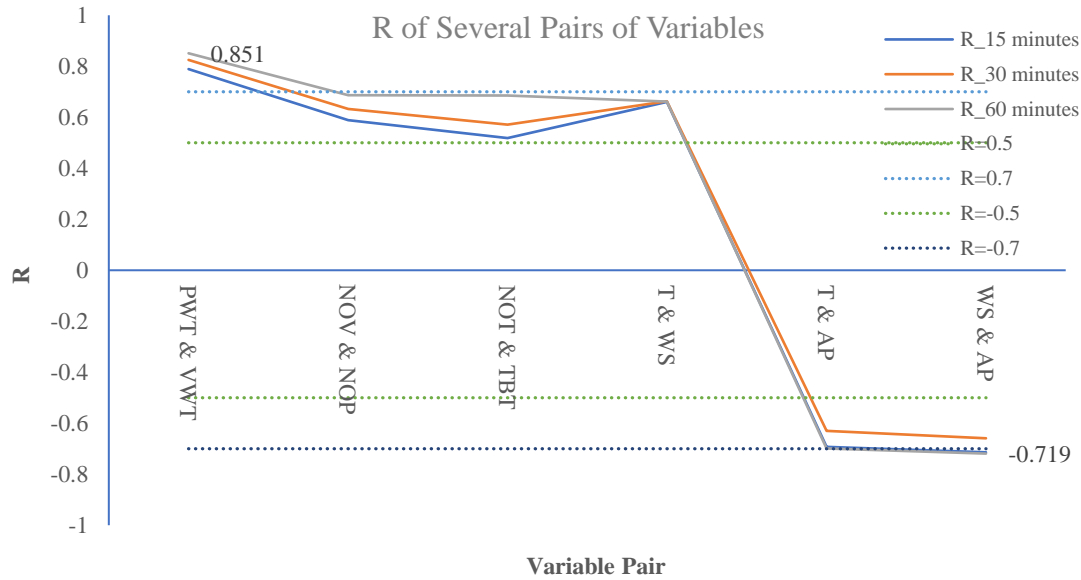


Figure 16

R Values of Several Pairs of Variables that Have at Least Moderate Correlation Relationship

The previous study (Zhao et al., 2019) demonstrated that traffic forecasting has always been a challenging task due to its complex spatial and temporal dependencies. Several traffic forecasting methods can use when just considering temporal dependence. For instance, respectively, they are Autoregressive Integrated Moving Average (ARIMA) Model (Ahmed et al., 1979; Hamed et al., 1995), the Kalman Filtering Model (Okutani et al., 1984), the Support Vector Regression Machine Model (Wu et al., 2004; Tao et al., 2006), the k-Nearest Neighbor Model (Zhang et al., 2009), and the Bayesian Model (Sun et al., 2006), and Partial Neural Network Model (Huang et al. 2014; Fu et al., 2017). To better make certain how simulation model inputs distribute in chronological order before simulating traffic volume within this study, analysis of temporal and trend distribution for each contribution is also supposed to carry out.

3. Temporal and Trend Distribution Characteristics of Traffic Variables

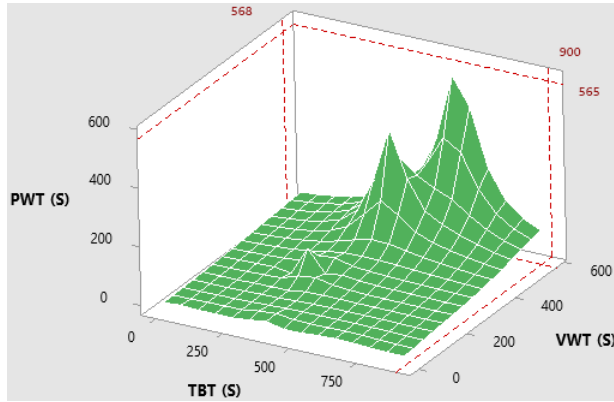
As it's known to all that before pedestrians or vehicles cross a specific area, they would evaluate how accessible the path is or how long it takes to pass through. Therefore, it seems

applicable to improving the condition of the sidewalk or road and upgrading the traffic signal system to deal with traffic issues such as congestion and blocking occur.

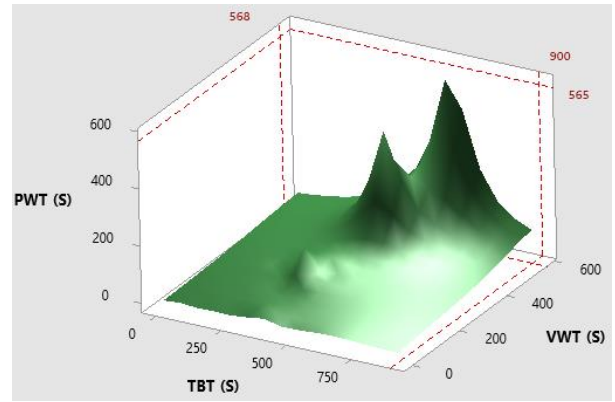
In this study, surface plots were drawn among NOP, NOV, and NOT during three different time intervals to determine the interconnection hidden behind these three variables. As shown in Figure 17 (a)- (b), when using every 15 minutes as the timespan, it happened that train blocked the way for pedestrians and vehicles up till to even a maximum of 900 seconds (15 minutes). In detail, pedestrians had to wait even up to 568 seconds (approximately 9.5 minutes) when the train blocked the access for 250 – 750 seconds (4 – 12 minutes). Similarly, demonstrated in Figure 18, train blocking time reached a peak of 1800 seconds (precisely 30 minutes) when choosing every 30 minutes as the time interval. However, there was no variation with the peak value of PWT, but a slight increment of 37 seconds with VWT despite timespan doubled from 15 to 30 minutes. Mostly, train blocking delayed vehicles and pedestrians approximately 200 – 600 seconds (3.5 – 10 minutes), and 200 – 565 seconds (3.5 – 9.25 minutes), respectively. Figure 19 revealed peak values of PWT and VWT remained the same as what they found in 30-min one, but TBT was almost increased in double from 1800 seconds (30 minutes) to 2365 (39.4 minutes). Within every 60 minutes, vehicles and pedestrians approximately delayed by 200 – 705 seconds (3.5 – 12 minutes) and 200 – 565 seconds (3.5 – 9.25 minutes), respectively, when train passing or standing at the cross.

Overall, no cases found that neither pedestrians nor vehicles waited a long time, which is equivalent to the length of each designated duration for access at the road intersection. Urgently, unacceptable delay of 39.4 minutes by cargo trains indeed had a significantly negative impact on residents' daily commute. Although pedestrians and vehicles can detour at nearby underpasses to

get a quick access crossing this periodically blocked intersection when train standing there, no matter how delays always exist.



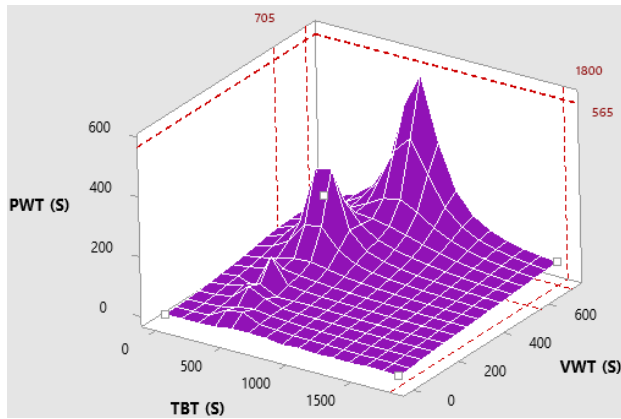
(a)



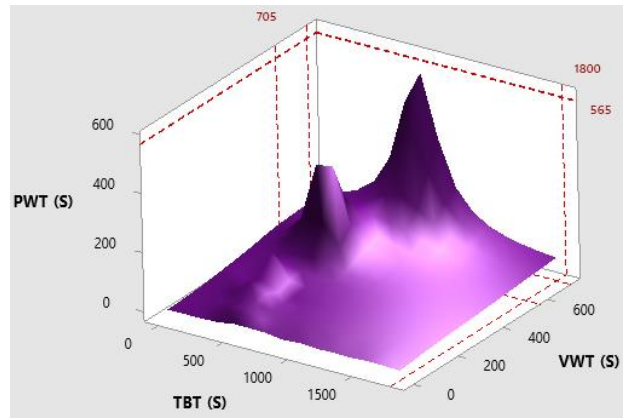
(b)

Figure 17 (a) – (b)

Surface Plot of PWT, VWT, and TBT (15 min)



(a)



(b)

Figure 18 (a) – (b)

Surface Plot of PWT, VWT, and TBT (30 min)

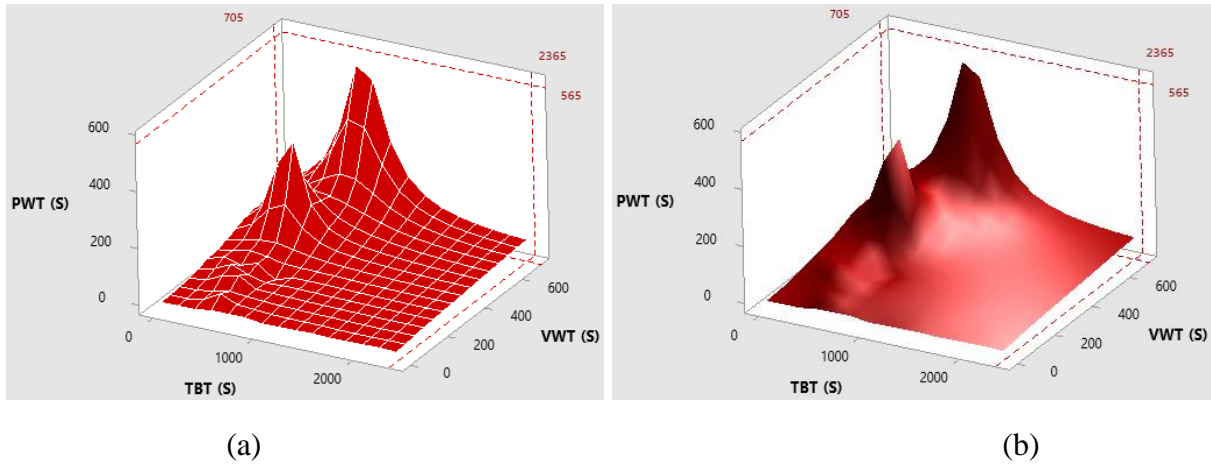


Figure 19 (a) – (b)

Surface Plot of PWT, VWT, and TBT (60 min)

Evidence from the original sampled videos also demonstrated that there was a high frequency of cargo train standing events. As shown in Table 14, long-time train blocking events happened again and again on October 20 and 23 of 2018, March 12 and 13 of 2019, respectively, of which caused delays by a minimum of 19 minutes and a maximum of 39.4 minutes.

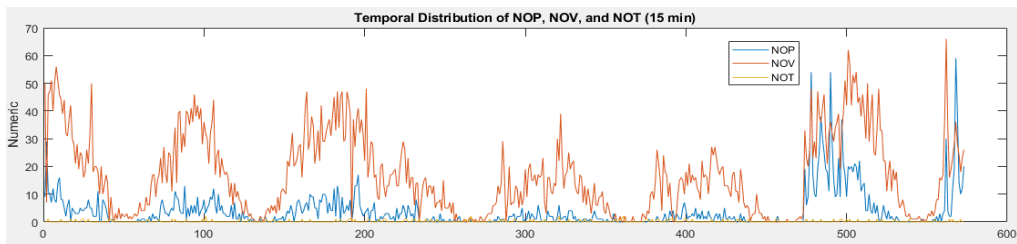
Table 14

Periodical Train Blocking Events

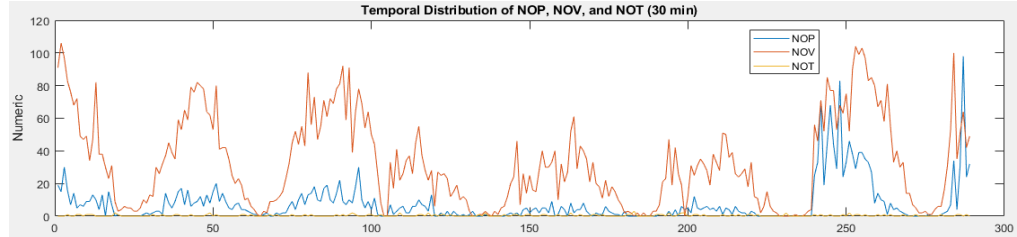
Date	Time	TBT (S)	TBT (Min)
10/20/2018	00:20 - 01:20	1746	29.1
10/23/2018	17:20 - 18:20	1964	32.7
3/12/2019	07:35 - 08:35	1405	23.4
3/12/2019	13:35 - 14:35	1140	19.0
3/13/2019	01:35 - 02:35	1715	28.6
3/13/2019	07:35 - 08:35	1016	16.9
3/13/2019	09:35 - 10:35	2365	39.4
3/13/2019	10:35 - 11:35	1119	18.7

To summarize the time-series distribution characteristics of traffic volume indexes, including NOP, NOV, and NOT, temporal distribution analysis for each of them conducted as demonstrated in Figures 20, 21, and 22, respectively. Roughly, in each unit of timespan designed

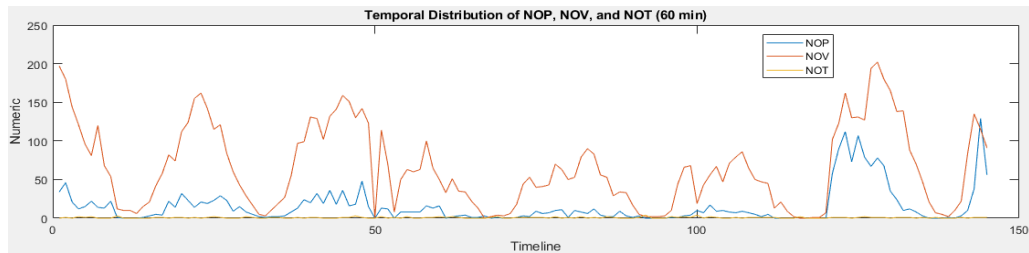
in this study, the volume of vehicles was almost always the greatest. The pedestrian volume took second place, and train volume was the last one. At the same time, both the pedestrian and vehicle volumes fluctuated like a waving under with passage of time. For daily variations of NOP and NOV, both had the same regular tendency that the sizes increase at the first stage then decrease. Besides, there were still some similarities that the peak values appeared at noon, and the trough values occurred at midnight. However, since trains ran strictly following the schedule, the volume of trains was never greater than four per unit time interval.



(a)



(b)

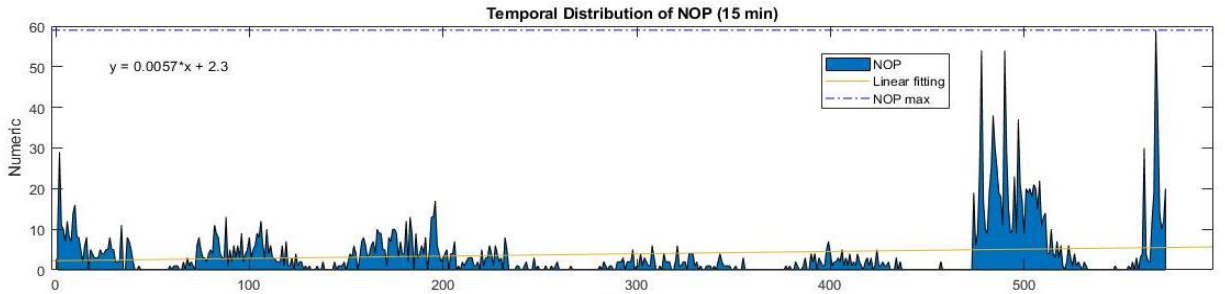


(c)

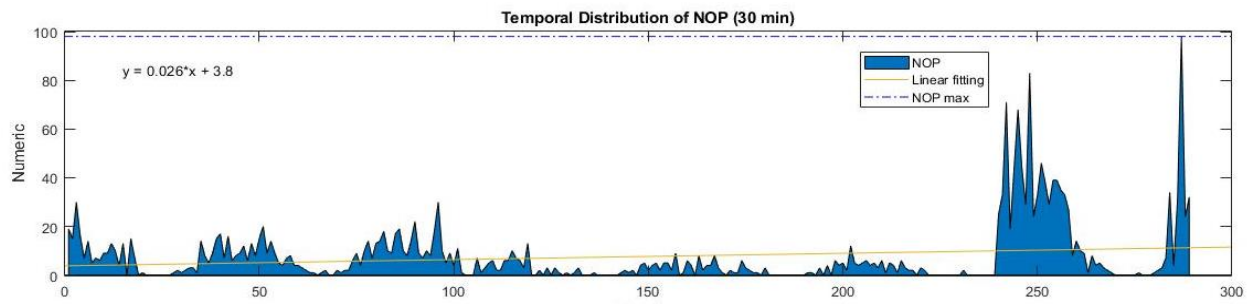
Figure 20 (a) – (c)

Temporal Distribution of NOP, NOV, and NOT (15, 30, and 60 minutes, respectively)

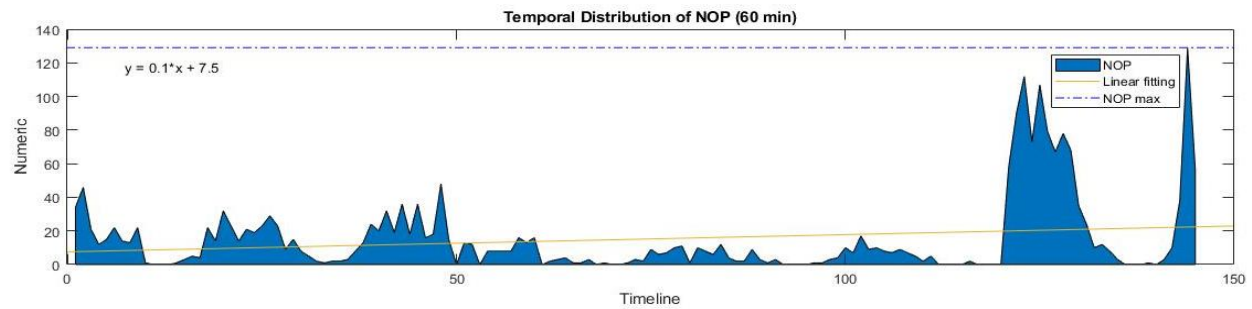
Through Figure 21 to 23, it's shown that, temporally, NOP had a slightly upward trend, but NOV did conversely when a linear fitting implemented. Meanwhile, results also indicated that when time interval increased from 15 to 30, then to 60 minutes, the absolute slope values of both fitting lines went up correspondingly. As expected, due to the scheduled operation of trains, there was a flat trend with the volume of trains.



(a)



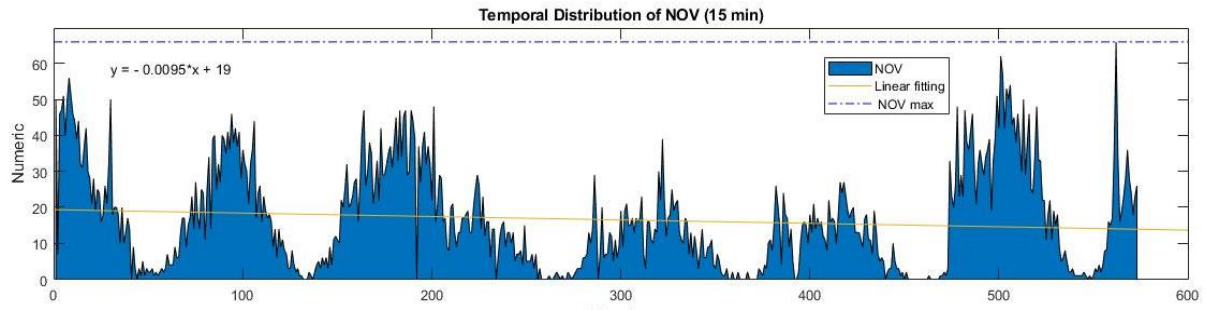
(b)



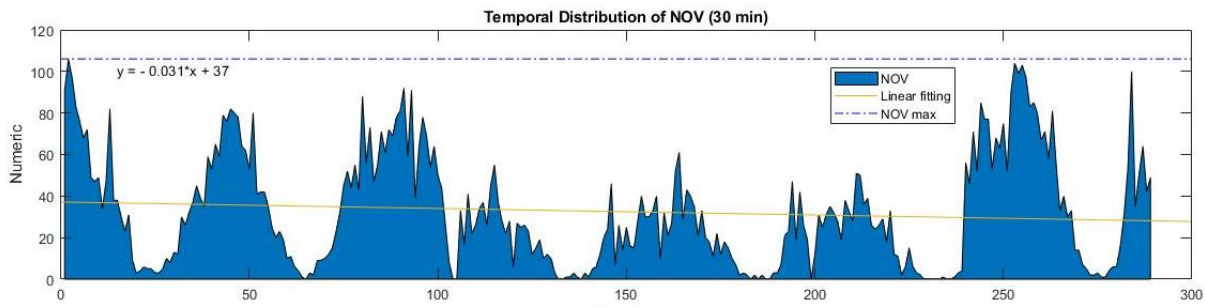
(c)

Figure 21 (a)-(c)

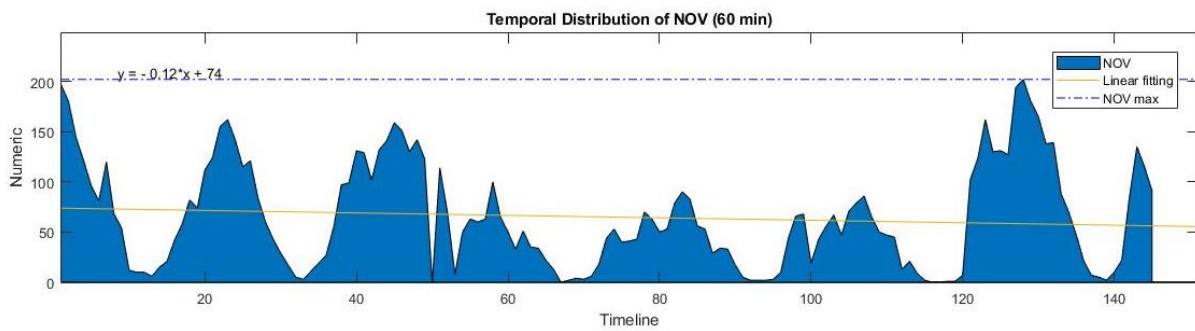
Temporal Distribution of Pedestrians (15, 30, and 60 minutes, respectively)



(a)



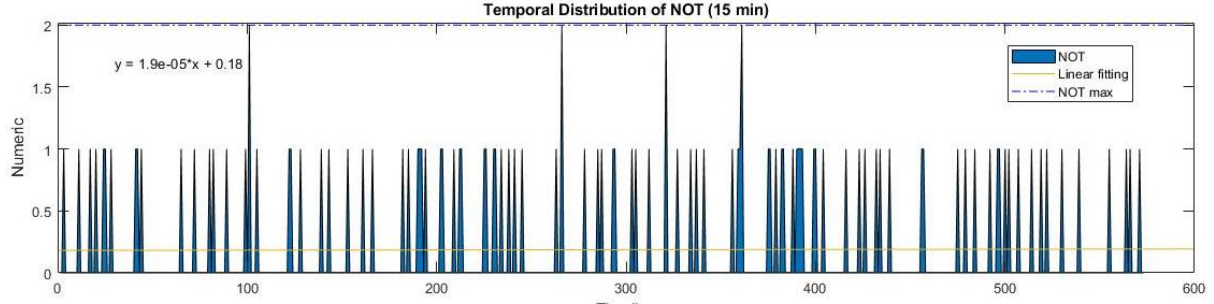
(b)



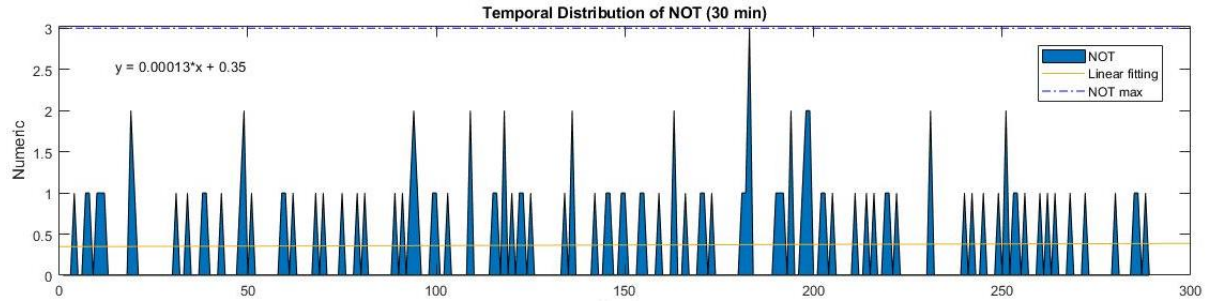
(c)

Figure 22 (a) – (c)

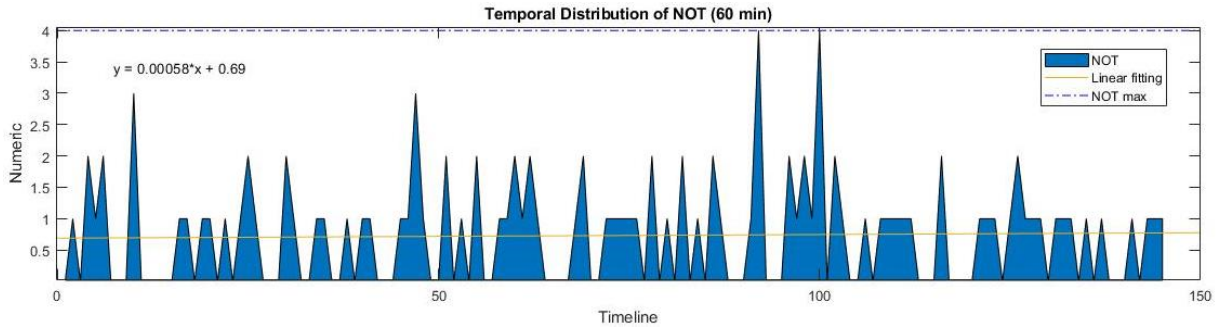
Temporal Distribution of Vehicles (15, 30, and 60 minutes, respectively)



(a)



(b)



(c)

Figure 23 (a) – (c)

Temporal Distribution of Trains (15, 30, and 60 minutes, respectively)

In summary, the results of the analytics of traffic and weather data indicated the following:

- 1) Pedestrians and vehicles had been blocked at the selected road intersection for up till to 565 and 705 seconds (9.4 and 11.7 minutes), respectively, by the periodically

standing cargo trains during 10/19/2018 – 10/20/2018, 10/22/2018 – 10/24/2018, 03/11/2019 – 03/14/2019, and 03/18/2019 – 03/19/2019.

- 2) Among those daily periods mentioned above, the maximum cargo train blocking time found to be 2365 seconds (39.4 minutes), which significantly affected the trafficability over the designated cross.
- 3) The volumes of both pedestrians and vehicles fluctuated like a wave with time. Meanwhile, both had the same daily tendency that the sizes increase at the early stage and then decrease, with peak values at noon and trough values at midnight.
- 4) NOP had a slightly upward trend, but NOV did conversely, as time went on. However, in each unit time interval, NOV was almost always higher than NOP. As expected, NOT was invariably the least all the time.

In the end, the consequences of statistical analysis also verified that linear-based machine learning algorithms are not applicable in this study due to the tremendous low strength of correlation relationships among most of the 11 variables. In other words, it intensively proved that the proposed ANN traffic simulation method has more benefits. The detail procedures of conducting ANN simulation will introduce in-depth within chapter 5. Consequently, the periodic long-time standing of cargo trains caused significant inconvenience to the residents, especially those who live around the pointed cross. Engineering measures such as building underpass or overpass for pedestrians and tunnel for both vehicles and pedestrians can consider.

Survey Data Analysis

In Europe, Vassi and Vlastos (2014) researched dealing with the categorization of cycling infrastructure across Europe and its contribution to the mobility management of cities with the implementation of an online questionnaire for data collection. A recent case (Ministry of

Transportation and Infrastructure of Canada, 2019) showed that an online survey has a strong capacity to collect people's feedback on the services such as roads, highways, and ports provided by the governmental agency mentioned when considering civil engagement. In the meantime, some key themes that matter regarding transportation include traffic management, overall condition of the pavement, highway improvements, etc. Similarly, in 2012, an online questionnaire was also executed as a partial data source for a feasibility study of a street improvement plan in Bloomington-Normal, Illinois. Following the previous studies listed, it found reasonable that implementing an online questionnaire to collect voices of residents, especially the potential service users. The analytical results of the online survey in this study discussed in more detail below.

Overall, 521 completed responses obtained from the online survey, with a total of 567 responses. A piece of Google map covering a practical size of 2-by-1.65 miles rectangular area divided into nine districts to roughly locate where the 567 participants of the online questionnaire live. Additionally, this map includes most of the urban area of the Township of Normal and partial northern metropolitan area of the City of Bloomington. Specifically, the research road cross is precisely the targeted spot shown in district no. 5 in Figure 24. In Figure 25, 13.2% of residents live close to the targeted intersection. Although there are more than 44% of participants who live in the area are not showing on the map. In other words, a slight change with transportation such as road renovation and new construction of pedestrian facilities would have a significant impact on their daily life.

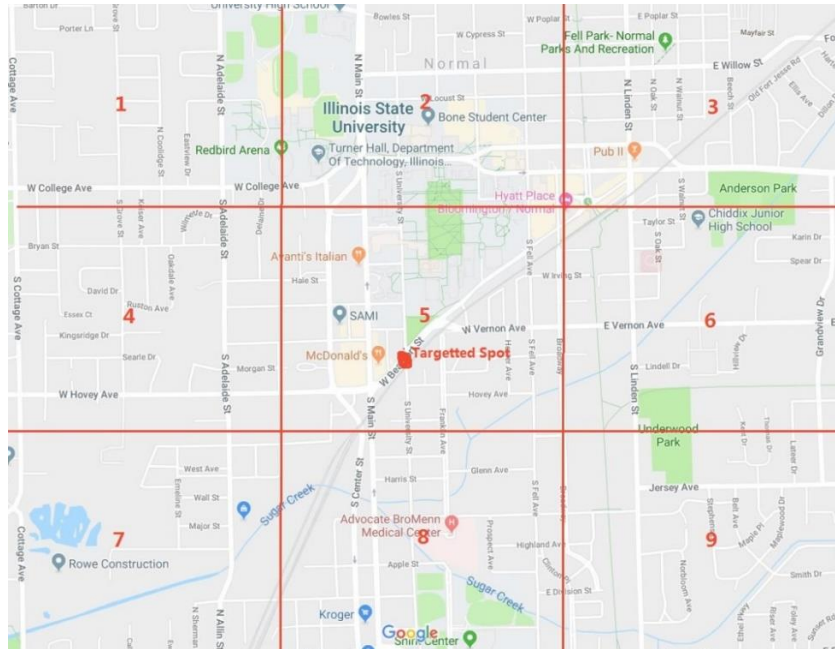


Figure 24

Residence Location of Participants in Map

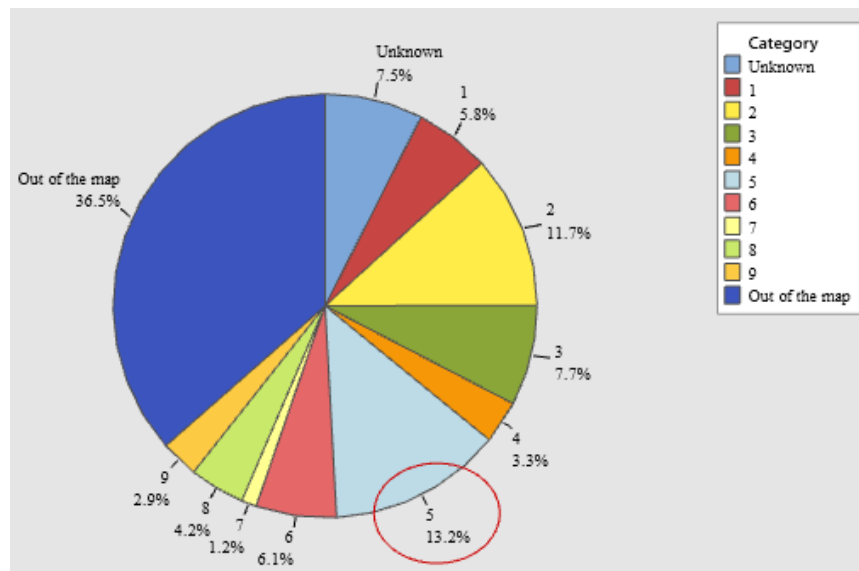


Figure 25

Pie Chart of Residence Location of Survey Participants

According to the questionnaire responses regarding the frequency of being blocked by trains weekly, unusually long cargo trains, as shown in Figure 26 (a), indicating that most of

them, precisely 38.5% of people, have ever been delayed with an interval 0 – 5 times. Then, 25.7%, 15.5%, and 5.3% of residents had to wait for 6 – 10, 11 – 15, and 16 – 20 times, respectively, to cross intersections of rail tracks and other roads in Bloomington-Normal (B-N) area. Surprisingly, 12.7% of people had the experiences of being blocked more than 20 times per week, compared with the other four groups. In Figure 26 (b), information can excavate from the shown bell-shape histogram that approximates half of the B-N residents delayed 6 – 10 minutes when the train blocked their access. With an increment of 5 from 6 to 20 minutes each time, the percentages of the population who had suffered train blocking decreased drastically from 46.5% to 22.5%, then down to 5.6%. Meanwhile, based on the practical experiences of residents' daily crossing at different intersections, 2.2% of them indeed faced being held up for more than 20 minutes (1200 seconds), which actively corresponded to the long-term cargo train blocking events detected in traffic videos in this study.

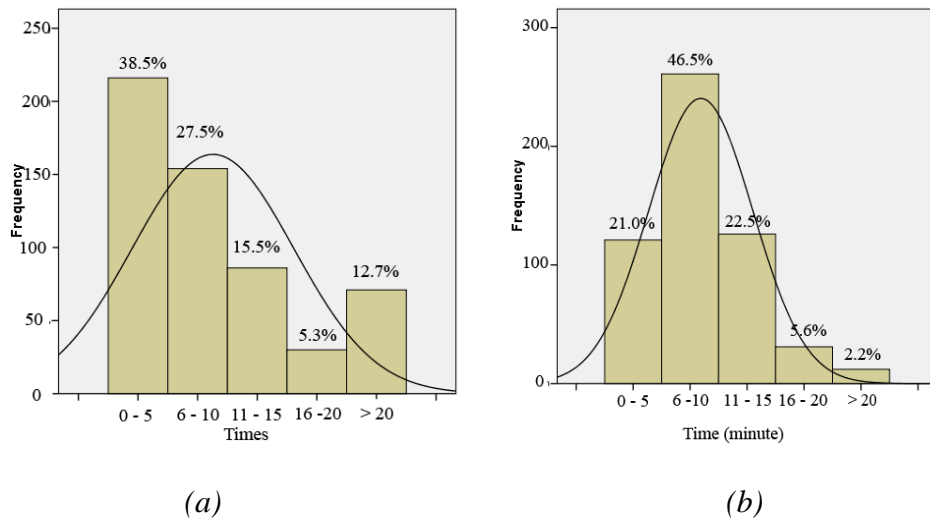


Figure 26 (a) – (b)

Train Blocking Frequency and Block Time

As far as the engineering measures that can take to lower down train blocking frequency and time, in Figure 28, four types of bridge, including cable-stayed bridge, truss, suspension, and

arch bridge, were given in the questionnaire for a ballot. Results, shown in Figure 28, revealed that the truss bridge was rated the most interesting one with 61.4% supporters. Then, more than ¼ liked arch bridge more. However, only 3.5% and 3.6% of people preferred cable-stayed and suspension bridges, respectively. Since the principal investigator initially categorized these bridges for the questionnaire, inevitably, some other possible designs such as pedestrian underpass or road tunnel hadn't been taken into consideration. Because vehicles were also substantially under the negative influence of the periodically standing cargo trains at the selected intersection under the findings shown in Figures 20 & 22. Therefore, to comprehensively conduct a feasibility study of transportation infrastructure projects, three kinds of structural designs consisting of pedestrian overpass, pedestrian underpass, and road tunnel should be considered in this study. For each of the mentioned architectural designs, multiple factors, including aesthetics, associated estimated budget, project schedule, potential social, economic, and environmental impacts, maintenance, risks, etc. are supposed to be analyzed in detail.

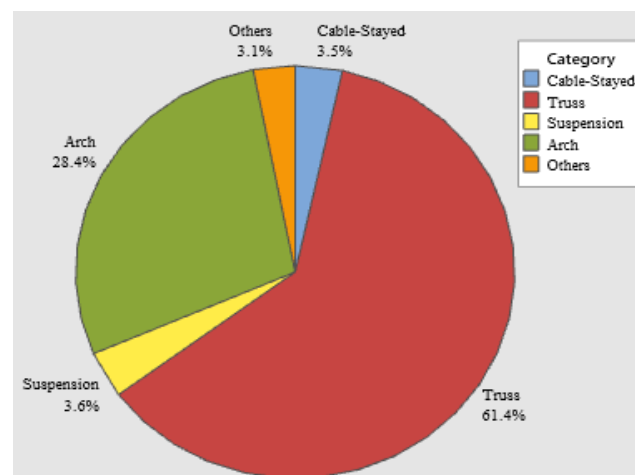


Figure 27

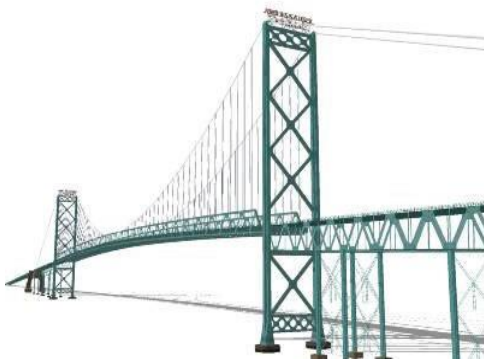
Ballots for Four Types of Bridge



(a) Cable-stayed Bridge



(b) Truss Bridge



(c) Suspension Bridge



(d) Arch Bridge

Figure 28 (a)-(d)

Four Types of Pedestrian Bridge

Taking one aspect of social impacts as an example, in terms of the benefits the proposed pedestrian bridge may have, shown in Figure 29, safety improvement was the most accepted one with 34% supporters. Accordingly, there is an actual demonstration from the public that eliminating the safety hazards for pedestrians and bicyclists at intersections of rail tracks and other roads indeed are the work priority of priorities. Meanwhile, 28% and 27% of voices showed an indication that a pedestrian facility like overpass would be beneficial to decreasing the traffic delays and enhancing the traffic mobility correspondingly. Another nonnegligible element pointed out that approximately 10% of residents thought a pedestrian bridge would

stimulate local economic development by setting it up as an iconic architecture. Based on the social impacts the proposed construction potentially has, we can conclude that social impact evaluation for any scheme of pedestrian overpass, underpass, and road tunnel ought to define for final decision-making in the feasibility study quantitatively.

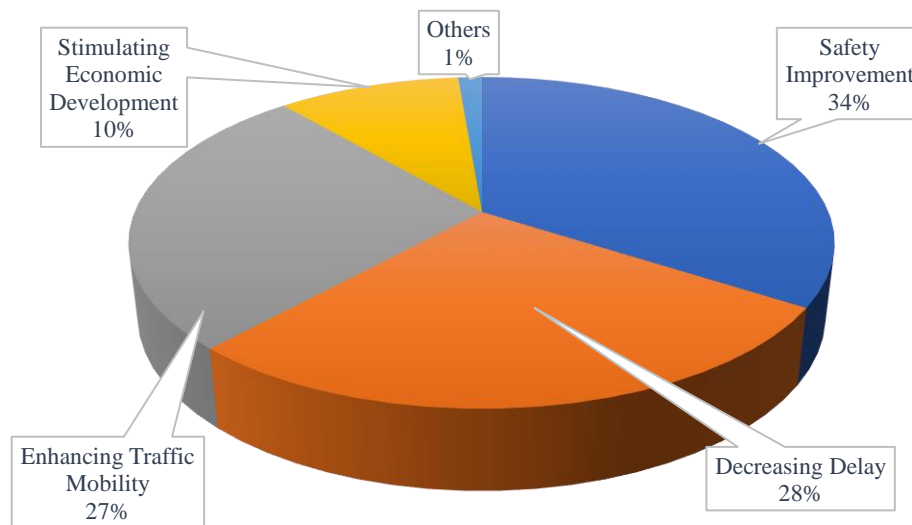


Figure 29

Social Impacts of the Proposed Pedestrian Overpass

When taking environmental sustainability into account, more than 1/3 of the public thought that utilizing eco-friendly, resilient materials and sustainable design can reduce the maintenance cost in the future. At the same time, installing a solar-panel roof on the bridge for generating and providing electricity for night lighting and network surveillance equipment was in favor of 35% residents. With a more convenient pedestrian facility, 29% of people regarded the overpass as advocacy that reducing the carbon emission by walking more instead of driving.

However, technically, under the National Environmental Policy Act of 1969 (NEPA, U.S.), a report of Environmental Impact Assessment (EIA) needs to create to identify and predict potential impacts. In detail, the environmental components such as water, soils, fauna, ecological

health, flora, and unique ecosystems should analyze minutely. Therefore, when evaluating the construction projects, the environmental impact should be necessarily taken into consideration. In the whole project lifecycle, an EIA will accompany, as shown in Figure 31.

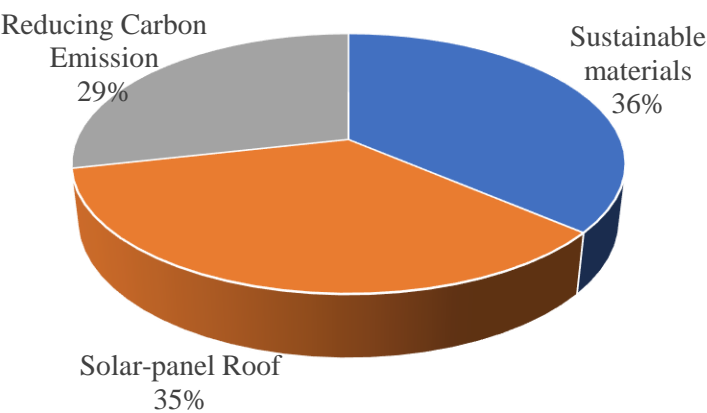


Figure 30
Sustainable and Environmental Impacts of the Pedestrian Bridge

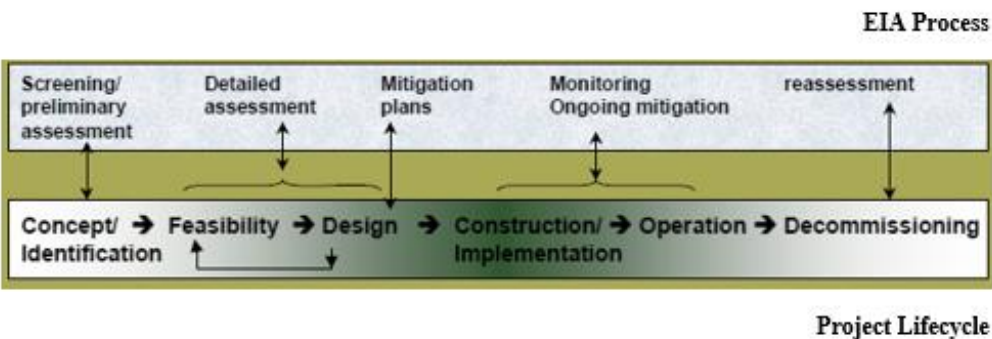


Figure 31
EIA Process and Project Lifecycle

CHAPTER V: MODEL SIMULATION AND VALIDATION

In this study, as demonstrated, MATLAB R2018b is the computer programming-based software applied for conducting traffic flow simulation for both pedestrian flow and vehicle flow. Afterward, this tool also plays a significant role in manipulating the fuzzy comprehensive evaluation for optimal selection of project schemes following fuzzy logic theory. In sequential order, traffic flow simulation associated with this study and thorough fuzzy evaluation of proposed construction schemes discussed in detail below.

Traffic Flow Simulation and Sensitivity Analysis

Principles of ANN Traffic Simulation in MATLAB

For the sake of programming BP neural network in the MATLAB environment, the following necessary but not limited to 10 steps constitute the main structure of the neural network, as shown in Figure 32. In this first place, datasets that consist of all the potential variables should be manually typed into the editor interface of MATLAB or saved with a format of the numeric matrix into a designated folder in MATLAB.

Previous studies (Sola et al., 1997) indicated that, concerning data normalization, the main goal is to scale data so that all data fall into a small and specific interval, such as (0,1) or (-1, 1). Beyond that, the other purpose dedicated to transfer the dimensional expression to a non-dimensional one. The benefits of this manipulation include improving the convergency rate and accuracy of modeling. Precisely, the commonly used methods such as linear or Min-Max, logarithmic, and inverse-cotangent normalization for this operation shows in Table 15.

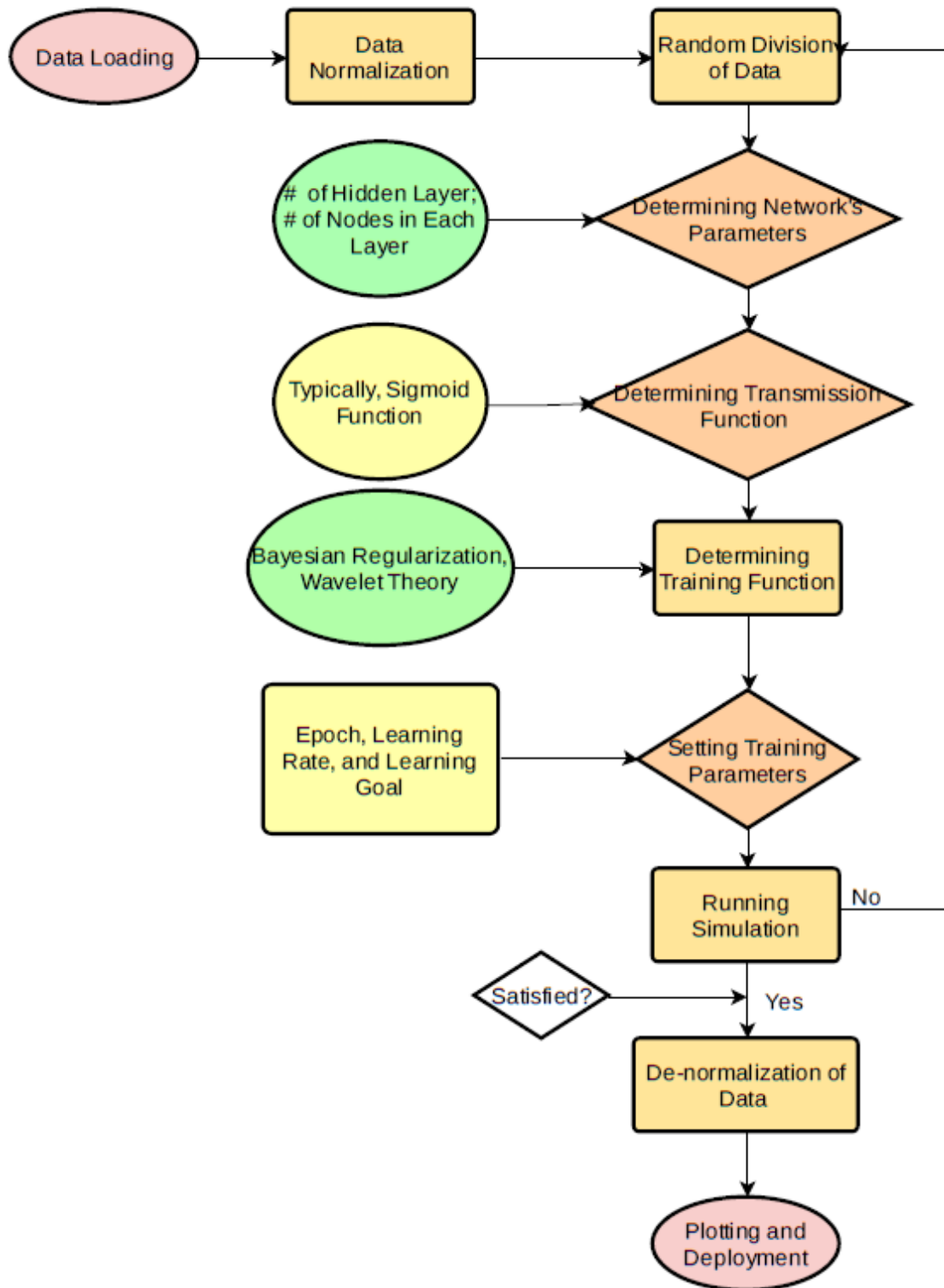


Figure 32

Application of ANN in MATLAB R2018b

Table 15*Commonly Used Ways of Normalization*

Methods	Mathematical Expression	References
Linear Normalization	$y = \frac{x - Min}{Max - Min}$	Jin et al., 2015
Logarithmic Normalization	$y = \log_{10}(x)$	Zavadskas, 2008 & Feng et al., 2014
Inverse Cotangent Normalization	$y = [\text{actan}(x)] \cdot 2/\pi$	Swanson, 2017

In the above table, x and y respectively represent the original and normalized data values.

In the given data array of each variable, Min is the minimum value, and Max is the maximum one. Relevant coding commands listed in Table 16 to achieve all three methods demonstrated in Table 15. In this study, ‘mapminmax’ was applied for composing the architecture of the neural network to conduct time-series BP neural network traffic flow simulation.

Table 16*Coding Commands for the Three Normalization in MATLAB*

MATLAB Commands	Scaling Interval
premnmx, postmnmx, and trmnmx	(-1,1)
prestd, poststd, and trstd	(0, 1)
mapminmax	(-1, 1)
mapstd	(0, 1)

In terms of random division of data, apart from additional testing of the proposed time-series BP neural network model, a specified proportion of input data will be picked up for training, validating, and testing model, respectively, when building up the model architecture. In detail, the majority of data will use for training the model, and the rest of the data are equally divided for the implementation of model validation and testing. Additionally, any of the input data points has the probability of being selected as one element in the training, validation, or testing dataset due to the randomness. In the end, to test the applicability of the proposed model,

additional testing can also be conducted. As shown in Figure 33, parameters, such as the number of input layer nodes, the number of hidden layers, and the corresponding number of nodes in each hidden layer, and the nodes of the output layer are supposed to determine in the process.

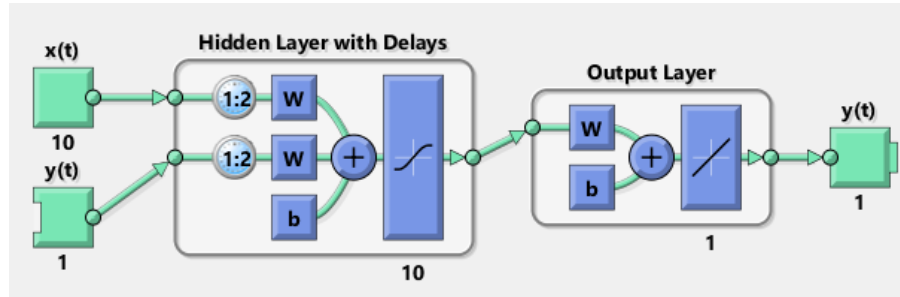


Figure 33

Basic Layout of Time-Series Neural Network

There should be an applicable method that can connect the adjacent layers. Hence, the transmission (activation) function that links nodes in the prior layer with the ones in the subsequent layer should also be appropriately determined because any badness in the whole process may lead to a significant simulation error. Generally, the sigmoid function is the ideal one for transiting coefficients, including weights and offsets calculated from the previous layer to the next when applying BP neural network for traffic flow prediction (Polson et al., 2017) and driver's lane-changing simulation (Bi et al., 2016).

Since the raw traffic data are dynamical in chronological order, when conducting time-series BP neural network traffic simulation, training functions introduced must be strongly applicable to considering temporal factors. A study conducted by Sun et al. (2004) indicated that the Bayesian neural network is practicable and useful for traffic flow forecasting in the field of intelligent transportation systems, even with incomplete data. Two years later, Zheng et al. (2006) proposed an approach named Bayesian Combined Neural Network for short-term traffic flow prediction. Considering the chaotic characteristics of traffic flow, Li et al. (2016) also came

up with to integrate the Bayesian Theorem and neural network algorithm for time-series traffic flow simulation. In the study, it proved that this method has high effectiveness in term of accuracy and timeliness for the short-term traffic simulation. In this study, according to the previous research findings, Bayesian regularization was accordingly brought into the time-series BP neural network traffic flow simulation. Technically, when training the neural network model, regularization plays a vital role in preventing overfitting by constraining the parameters which should be optimized. To better understand Bayesian regularization, relevant example and explanations demonstrated in the following:

Generally, supervised learning can view as minimizing the objective equation (23):

$$w^* = \operatorname{argmin} \sum_i L(y_i, f(x_i; w)) + \mu \phi(w) \quad (23)$$

where, the first item $L(y_i, f(x_i; w))$ is the error between the predicted value, $f(x_i; w)$, of the i^{th} sample and the actual value of y_i . $\phi(w)$ is the regularization function of parameter w .

$\sum_i L(y_i, f(x_i; w))$ And $\mu \phi(w)$ is the training error and testing error, respectively.

Both items should minimize as much as possible to obtain the minimum error. Therefore, Bayesian Theorem aims at estimating the subjective probability of some unknown states under incomplete information and using the Bayesian formula to correct the probability of occurrence, then applying the expected value and modified probability to make the optimal decision.

Assume that events E_1, E_2, \dots, E_n are elements in sample set S ; $P(E_i)$ is the occurrence probability of event E_i , and $P(E_i) > 0$ ($i = 1, 2, \dots, n$). For any other event x , $P(x) > 0$, following Bayesian Theorem, we can conclude that:

$$P(E_j/x) = \frac{p(x/E_j)P(E_j)}{\sum_{i=1}^n P(x/D_i)P(D_i)} \quad (24)$$

Based on the given observation or model-establishing dataset E, Bayesian Theorem can be used to maximize the posterior probability of parameter w. Then, another equation can generate below.

$$w^* = \operatorname{argmax}_w P(w/E) = \operatorname{argmax}_w \frac{p((E/w) \cdot p(w))}{p(E)} = \operatorname{argmax}_w p(E/w)p(w) \quad (25)$$

where $P(w/D)$ is the likelihood function, which denotes the occurrence probability of dataset E when parameter w exists. $p(w)$ is the prior probability of parameter w.

For likelihood function of $P(w/D)$, it can also demonstrate like the following:

$$P(E/w) = \prod_{k=1}^n p(E_i/w) \quad (26)$$

Then, converting equation (16) by taking the logarithm, a new equation is shown below:

$$P(E/w) = \operatorname{argmin}_w - \sum_{i=1}^n \log p(E_i/w) - \log p(w) \quad (27)$$

When prior probability distributes regularly,

$$p(w_i) = N(w_i/\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w_i-\mu)^2}{2\sigma^2}} \quad (28)$$

Plug equation (18) into (17), a new equation (29) can achieve.

$$w^* = \operatorname{argmin}_w - \sum_{k=0}^n \log p(E_i/w) + \gamma \sum_{k=0}^m (w_i)^2, \quad (\mu = 0, \sigma = \sqrt{\frac{1}{\gamma}}) \quad (29)$$

The above detailly discussed the principle of Bayesian regularization, which acts as the training function. Moreover, it's also essential to subjectively set up training parameters such as the number of epoch (iteration), training goal, and training rate before running simulation. When the complete simulations and satisfying, de-normalization, the different steps of normalizing input values, must be executed to reverse the predicted values back to their original format for reading. Finally, the corresponding coefficient matrix comprising of the weights of nodes and offset in each layer will produce for future application and deployment.

With regards to the degree of satisfaction with the modeling, mean squared error (MSE) will be used to denote the average squared difference between the predicted values and actual or expected value. In theory, a smaller MSE refers to a lower error. In statistics, MSE can demonstrate in equation (30).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (X_i - \widehat{X}_i)^2 \quad (30)$$

where \mathbf{X} is the vector of actual or observed values of the variable being predicted, with $\widehat{\mathbf{X}}_i$ being the predicted values. The MSE is the mean ($\frac{1}{n} \sum_{i=1}^n$) of the squares of the errors $(X_i - \widehat{X}_i)^2$.

In the meantime, since additional testing will also conduct for verifying the performance of model prediction, beyond the MSE, correlation coefficient R is also selected as the quantitative indicator of the degree of relationship between actual and predicted values. The higher the R is, the better the additional test works, then the more applicable the model is.

Implementation of ANN Traffic Simulation in MATLAB

In this study, based on the results obtained from the previous statistical analysis of traffic video data, two variables, including NOP and NOV, were determined as the potentially influential factors. Concerning the possible social impact, these two affect the decision-making of implementing a construction project to deal with the safety hazards for pedestrians and improve the traffic mobility for vehicles as well. ANN Traffic simulation models simulate the 15-minute, 30-minute, and 60-minute dynamic flow of pedestrians and cars, respectively. Since traffic data randomly sampled, sampled datasets of real-time traffic dynamics and corresponding weather data ordered chronologically. In detail, traffic data mined from the video recorded during October 19-20 and October 22-24 of 2018, March 11-14 and March 18 of 2019, and corresponding meteorological data used to run the neural network traffic simulation model. To

further verify the performance of each model, data mined from the transportation videos recorded on March 19 of 2019 imported to run additional testing for model validation purposes.

To determine the optimal network's parameters such as the number of nodes in the hidden layer and the number of delays when conducting the time-series traffic flow neural network simulation, a series of trial modeling was executed based on the data of 15-min NOP. On the strength of the underlying neural network formula, the problem definition of time-series BP neural network traffic flow simulation with Bayesian regularization can demonstrate as equation (31) below.

$$y(t) = f(x(t-1), \dots, x(t-d), y(t-1), \dots, y(t-d)) \quad (31)$$

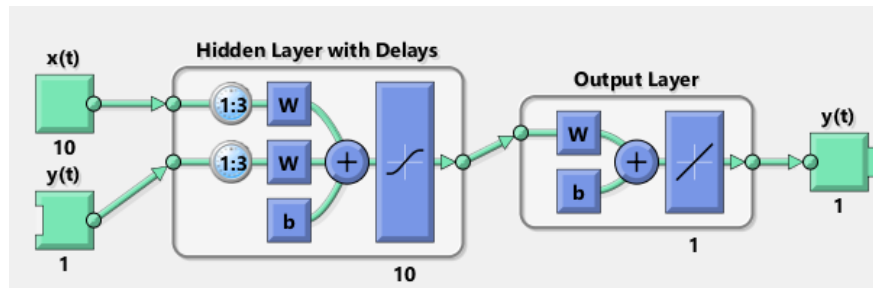
where, $y(t)$ is the current predicted value; $f(x, y)$ means the essential function of time-series BP neural network; $x(t-d)$ is the influential variable with a delay of d time units; $y(t-d)$ means the predicted value with a delay of d time units.

As shown in Table 17, there are 18 trials conducted. It found that trial simulation no. Five performed the best due to it had the most significant correlation coefficient R-value with 0.9515 of regression between predicted values and actual value, and the most acceptable MSE for both training and testing when establishing the model. Then, we can conclude that the number of nodes in the hidden layer and the number of delays can determine as 10 and 3 (shown in Figure 34). Meanwhile, randomly, 70%, 15%, and 15% of the observation or model-establishing data should be used to train, validate, and test the model, respectively. Then, equation (32) can specify as below:

$$y(t) = f(x(t-1), x(t-2), x(t-3), y(t-1), y(t-2), y(t-3)) \quad (32)$$

Table 17*Determination for Network's Parameters*

No.	Network's Parameters		Percentages of Data Division (%)			R Values of Regression			MSE	
	# of Nodes	# of Delays	Training	Validation	Testing	Training	Test	All	Training	Testing
1	11	3	60	20	20	0.9816	0.3328	0.8876		
2	11	3	70	15	15	0.9786	0.5406	0.8466		
3	11	3	80	10	10	0.9893	0.2246	0.8102		
4	10	3	60	20	20	0.9888	0.5329	0.8679		
5	10	3	70	15	15	0.9773	0.6856	0.9515	1.89	13.56
6	10	3	80	10	10	0.9708	0.4882	0.8943		
7	9	3	60	20	20	0.9753	0.4639	0.9004		
8	9	3	70	15	15	0.9801	0.6170	0.9180		
9	9	3	80	10	10	0.9805	0.5318	0.922		
10	11	2	60	20	20	0.9675	0.4018	0.8459		
11	11	2	70	15	15	0.9502	0.5238	0.8610		
12	11	2	80	10	10	0.9539	0.7451	0.9002	2.96	47.03
13	10	2	60	20	20	0.9670	0.3173	0.8555		
14	10	2	70	15	15	0.9416	0.6606	0.8548		
15	10	2	80	10	10	0.9614	0.5381	0.8489		
16	9	2	60	20	20	0.9525	0.5932	0.8628		
17	9	2	70	15	15	0.9798	0.8067	0.8826	1.49	73.12
18	9	2	80	10	10	0.9750	0.5274	0.9478		

**Figure 34***The Layout of the Most Optimal Time-Series Neural Network*

In total, six time-series BP neural network simulations were conducted based on the optimal network's parameters obtained from those 18 trials. To compare the accuracies of pedestrian flow and vehicle flow simulations and determine the optimal simulation model among

the three different unit time intervals. Most importantly, four types of indicators, including MSE-based model training and testing performance, error histogram, correlation-based (R) regressions between predicted values and observed ones, and the time-series response of every single modeling detailly discussed below. The following introduces the 15-min, 30-min, and 60-min pedestrian/vehicle flow simulation in MATLAB environment, respectively.

1. 15-minute pedestrian/vehicle flow simulation

As shown in Figure 35, it found to be that the training phase of the time-series neural network simulation model for pedestrian flow reached the best performance when running 176 iterations with an MSE of 1.9427. Concerning the errors between targets and outputs, they overall lay on an interval of (-5.8, 5.92). Besides, Figure 36 also indicated that most errors of the 15-min pedestrian flow simulation occurred at either -1.111 or 1.233.

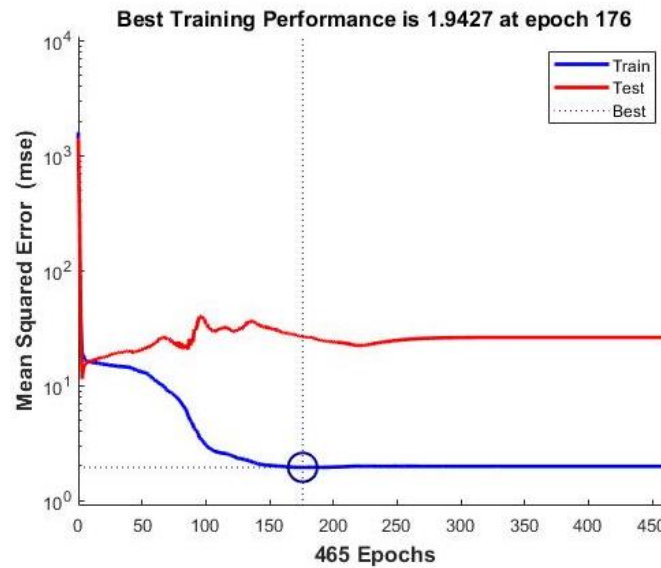


Figure 35

Training Performance of 15-Minute NOP Simulation

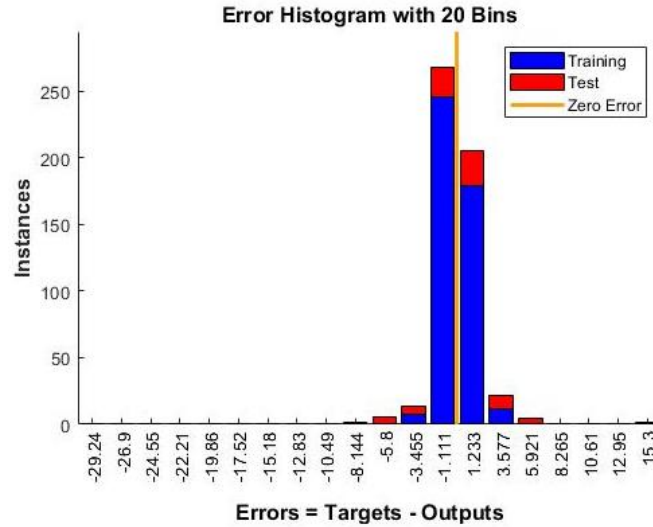


Figure 36

Error Histogram of 15-Minute NOP Simulation

In Figure 37 (a)-(c), correlation coefficient R was used to show the degree of correlation between targets and outputs. Since both training and testing sections automatically executed when running the model in MATLAB, there were three different correlation coefficients used to reflect the strength of correlation for each corresponding session and the overall modeling. Visually, the training session of the 15-min pedestrian flow simulation had a very high R-value of 0.9756. However, due to a relatively low R for the testing session, the overall R of this simulation is only 0.9283. Subsequently, in Figure 38, a time-series response of this 15-min pedestrian flow simulation is demonstrated. Overall, the model showed a significantly high accuracy on account of a few conspicuous errors that occurred at the end of the time series.

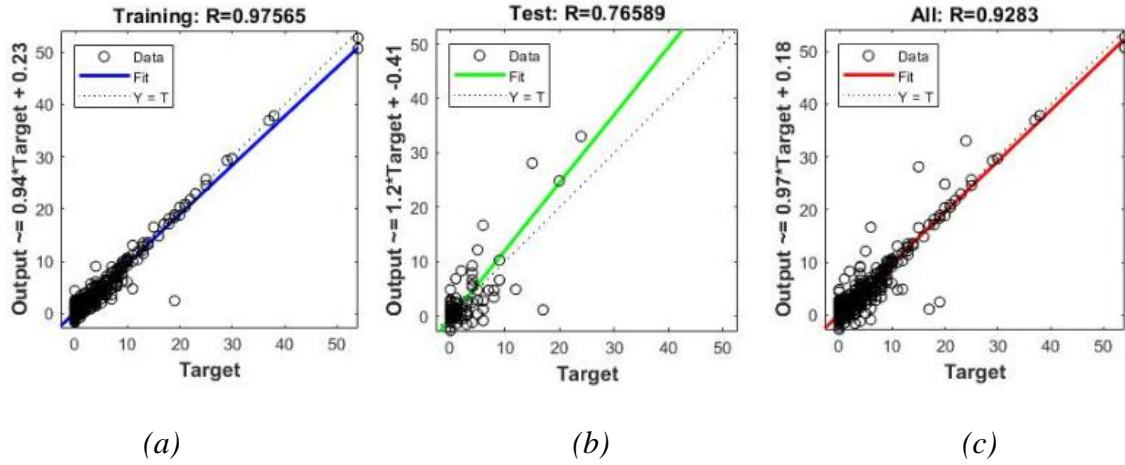


Figure 37 (a)-(c)

Regressions of 15-Minute NOP Simulation

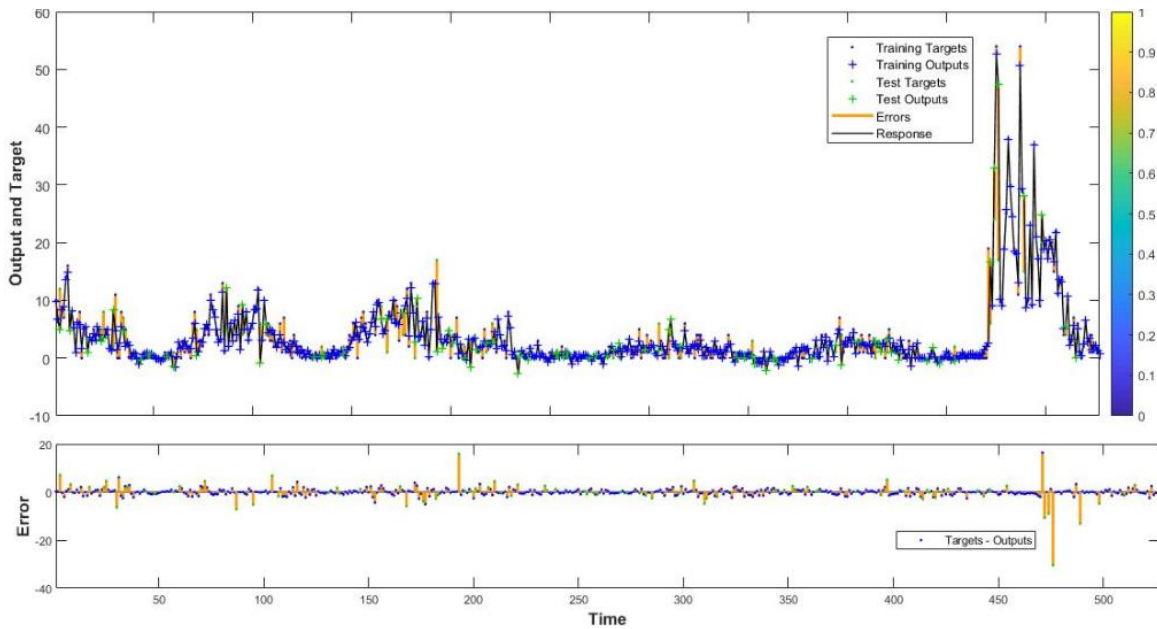


Figure 38

Time-Series Response and Error of 15-Minute NOP Simulation

For 15-min vehicle flow simulation, surprisingly, as demonstrated in Figure 39, the iteration even reached up to 1000 times. The best training performance appeared with an MSE of 22.68. According to Figure 40, however, compared with the simulation error of 15-min

pedestrian flow simulation, the 15-min vehicle one had a more extensive range from -18.43 to 28.90. Meanwhile, the error roughly distributed normally in a bell shape.

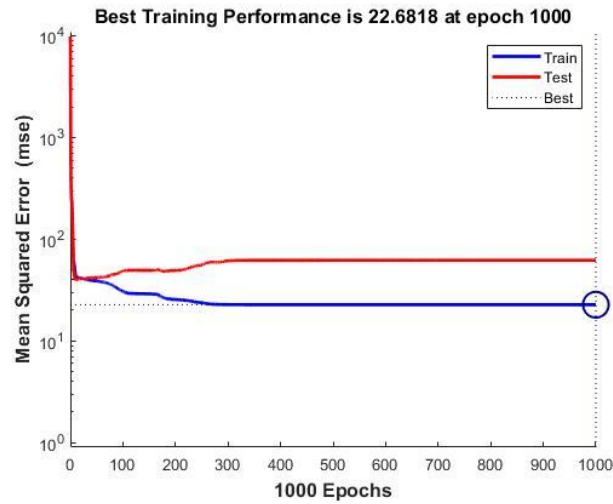


Figure 39

Training Performance of 15-Minute NOV Simulation

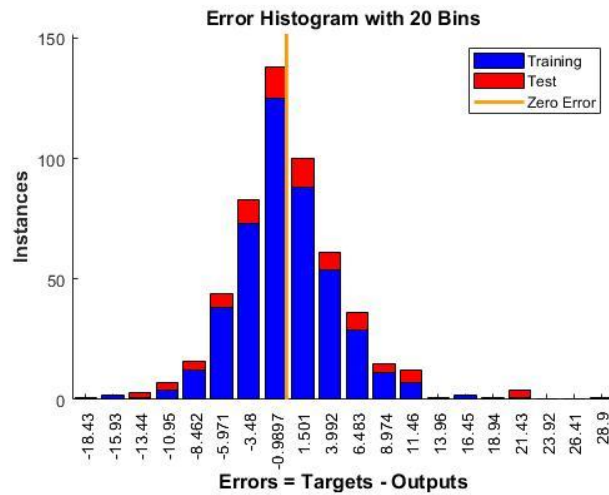


Figure 40

Error Histogram of 15-Minute NOV Simulation

In terms of correlation strength between predicted and observed vehicle flow, they are 0.9428, 0.8509, and 0.9267 for the training session, testing session, and the overall simulation,

respectively, shown in Figure 41 (a)-(c). Nevertheless, for the time-series response of 15-min vehicle flow simulation, in Figure 42, the profile looked rough due to significant errors with the whole series.

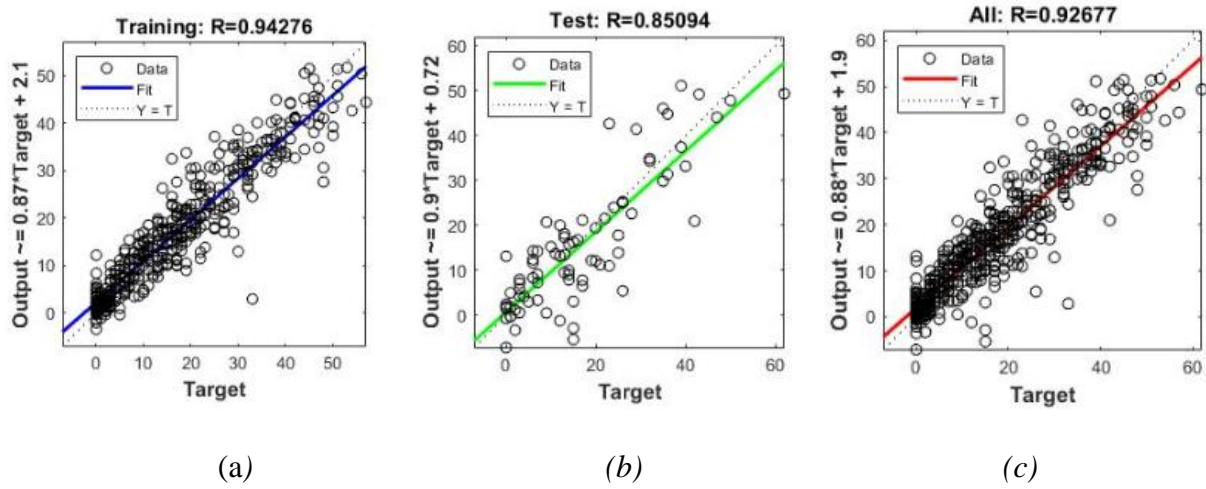


Figure 41(a)-(c)

Regressions of 15-Minute NOV Simulation

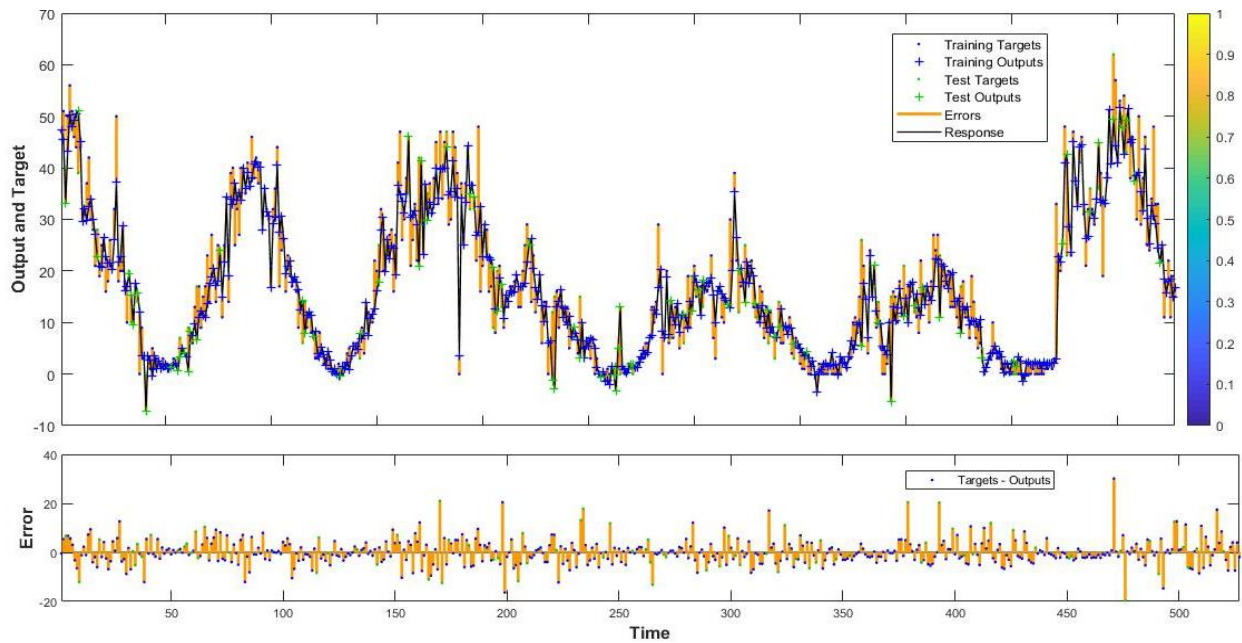


Figure 42

Time-Series Response of 15-Minute NOV Simulation

2. 30-minute pedestrian/vehicle flow simulation

When expanding the timespan from 15 to 30 minutes, all the model indicators changed correspondingly. For instance, the 30-min pedestrian flow simulation model achieved the best training performance at 411 iterations with an MSE of 3.30, as demonstrated in Figure 43.

With regards to the errors of the 30-min pedestrian flow simulation model, in Figure 44, it discovered that the error range $(-16.65, 13.97)$ was expanded drastically with a comparison of the one, $(-5.8, 5.92)$, of the 15-min pedestrian flow simulation. Based on this case, it can primarily prove that pedestrian flow simulation has more significant errors as the increment of the unit time interval.

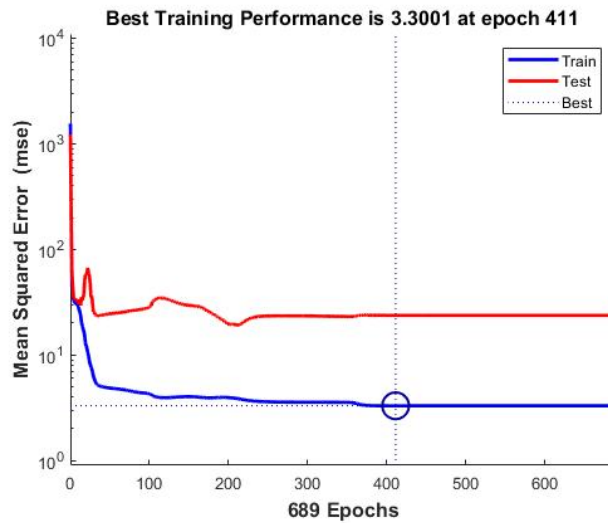


Figure 43

Training Performance of 30-Minute NOP Simulation

Although the errors increased significantly, however, the best goodness of fit with 30-min pedestrian flow simulation occurred, due to a training R of 0.9862, testing R of 0.9285, and an R of 0.9746 for the whole model, shown in Figure 45 (a)-(c).

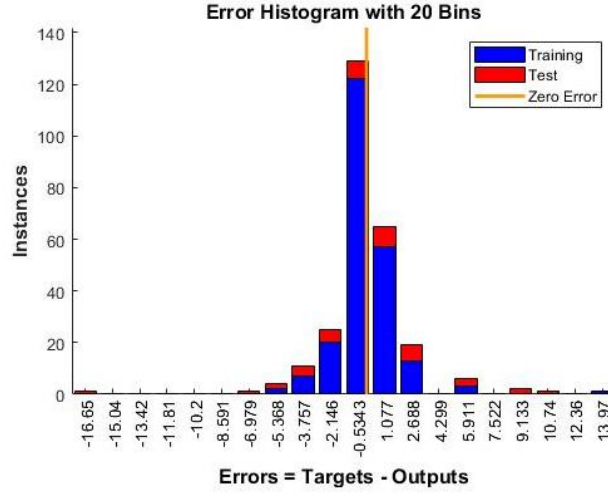


Figure 44

Error Histogram of 30-Minute NOP Simulation

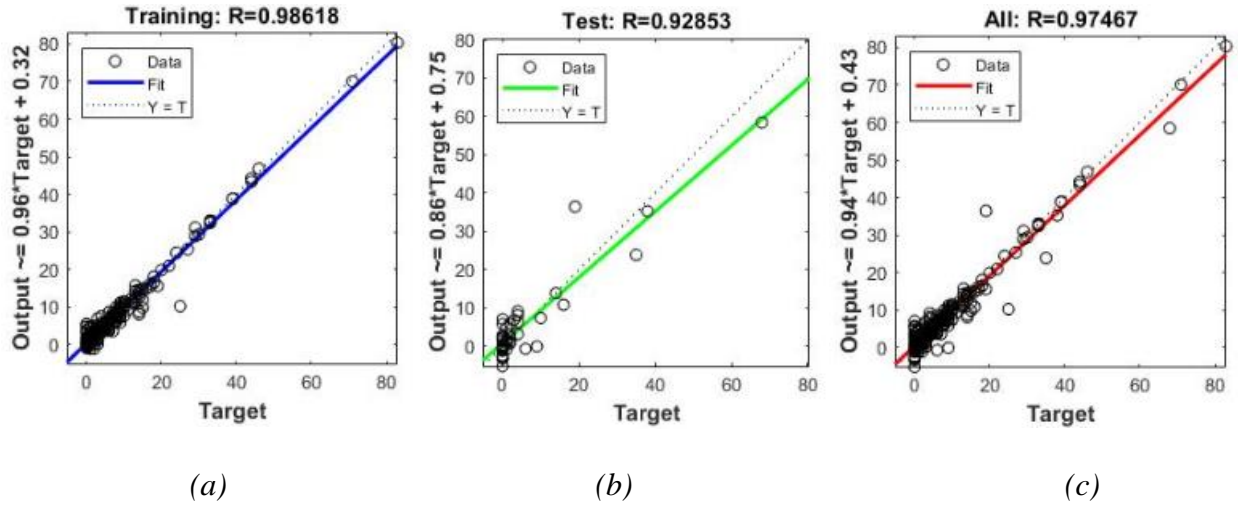


Figure 45 (a)-(c)

Regressions of 30-Minute NOP Simulation

In Figure 46, the time-series response curve approximates the target one because of a small number of variations between the targets and model outputs. Besides, the errors of 30-min pedestrian flow time-series simulation don't tend to be apparent apart from the ending section of the whole series. In other words, most of the errors lie in a range between -10 and 10, but the absolute value of the highest error is no more than 20.

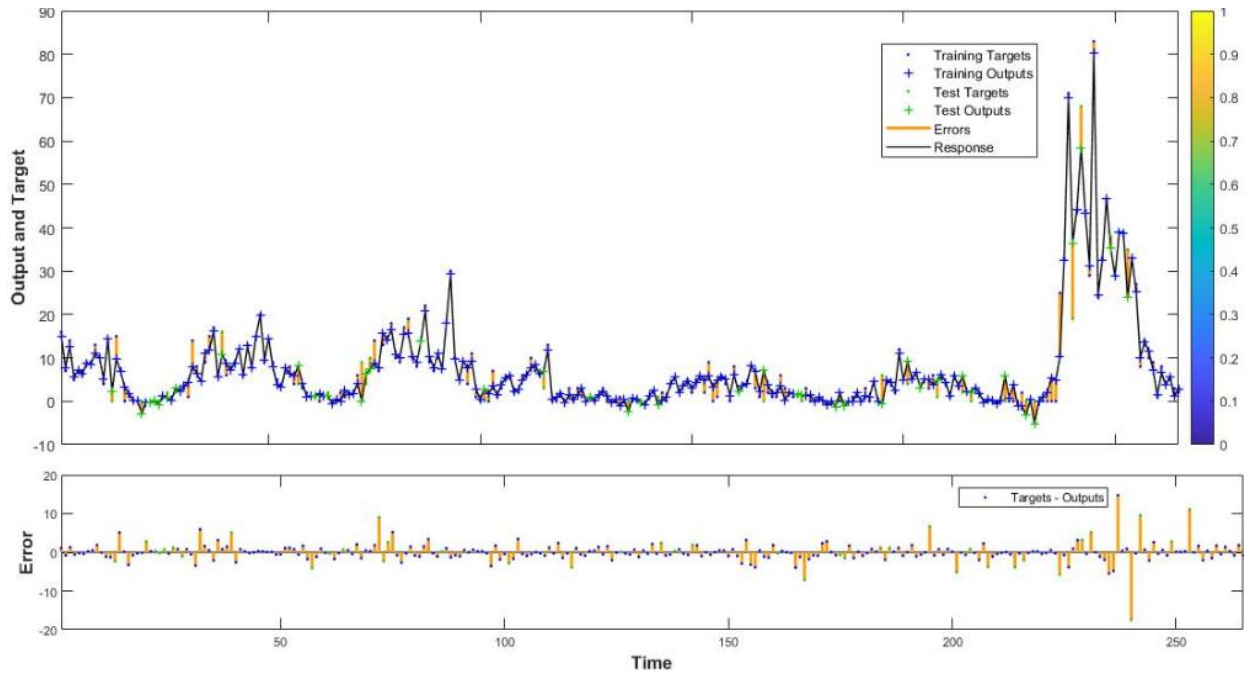


Figure 46

Time-Series Response of 30-Minute NOP Simulation

As for the 30-min vehicle flow simulation, Figure 47 revealed that there were 288 modeling iterations conducted before achieving the best model training performance with an MSE of 120.47 at the 67th epoch. Comparatively speaking, this model ran more efficiently than the 15-min vehicle flow one due to a smaller number of iterations but had an MSE, which was almost six times more than the one of the 15-min simulation. As shown in Figure 48, the training errors and testing errors varied within an interval of $(-25.77, 44.08)$, and $(-14.74, 40.40)$, respectively. Mostly, both the training and testing errors gathered at the range of $(-7.385, 7.319)$.

Intuitively, shown in Figure 49 (a)-(c), the testing R of 30-min NOV simulation is the greatest with the number of 0.9335. However, the data points, including targets and predicted ones, used for the regressions of both training sections and the whole model distributed relatively discretely although there was a training R of 0.9069 and an all R of 0.9107.

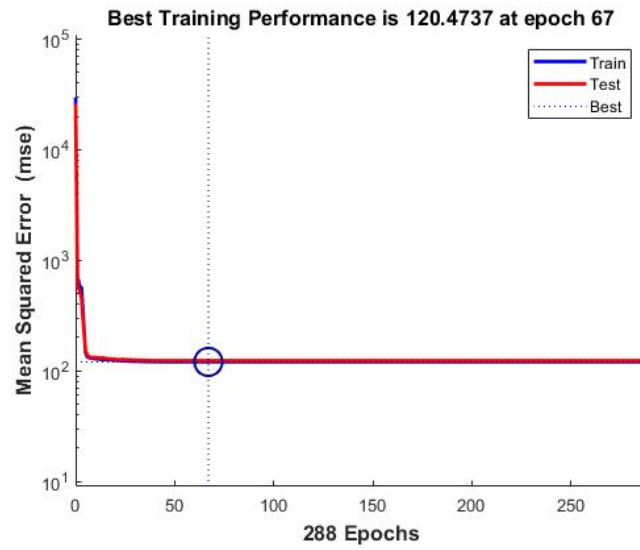


Figure 47

Training Performance of 30-min NOV Simulation

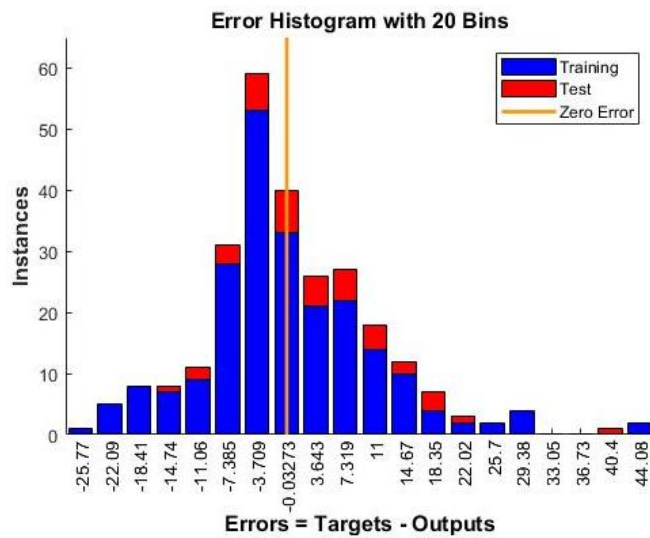


Figure 48

Error Histogram of 30-min NOV Simulation

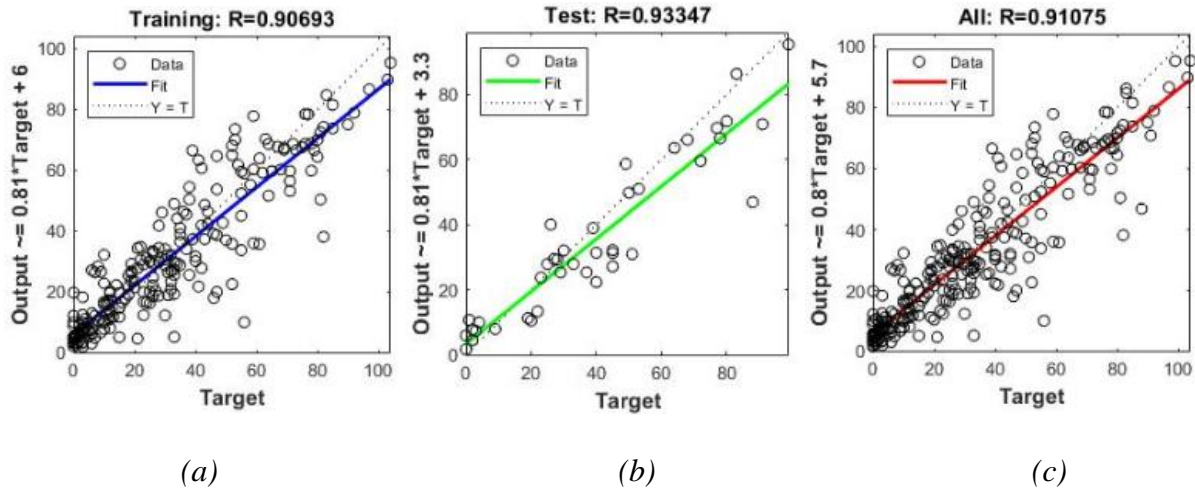


Figure 49 (a)-(c)

Regressions of 30-Minute NOV Simulation

In Figure 50, it's shown that the time-series response of the 30-min vehicle flow simulation had a smoother curve than the one of the 15-min simulation, but also generated more significant errors which lied in an interval of $(-40, 60)$.

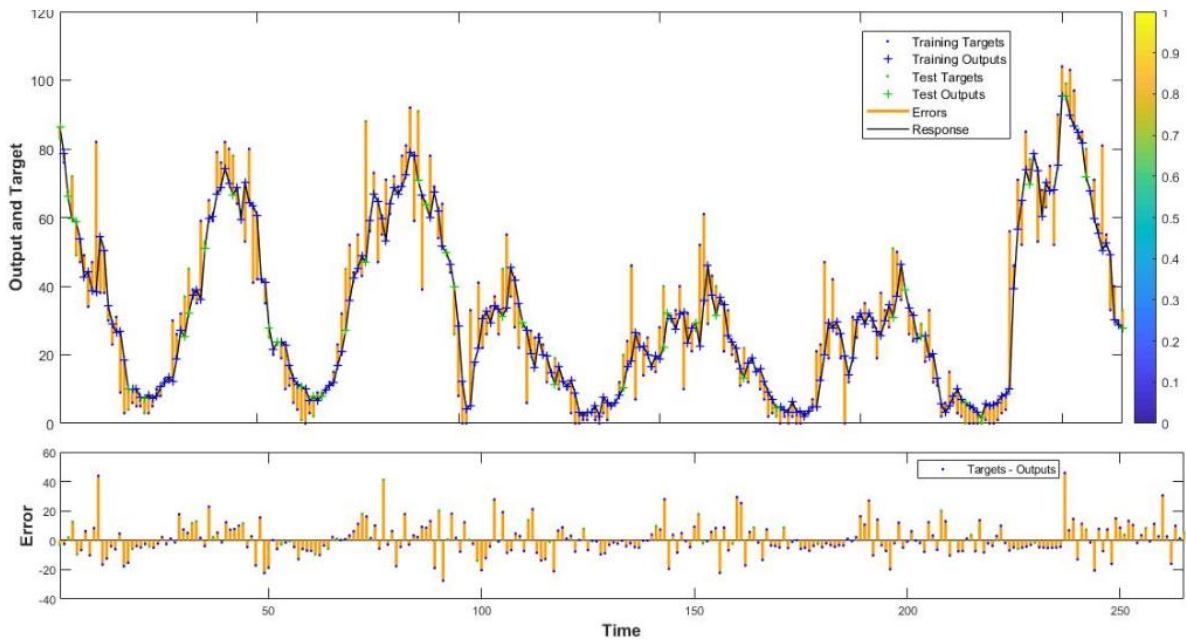


Figure 50

Time-Series Response of 30-Minute NOV Simulation

3. 60-minute pedestrian/vehicle flow simulation

When doubling the unit time interval from 30 minutes to 60 minutes (an hour), both the pedestrian and vehicle flow simulations dedicated some differences compared with either the 15-min or 30-min ones. For instance, taking 60-min pedestrian flow simulation as an example, as shown in Figure 51, the model training section reached the best performance at 251 epochs with an MSE of 21.69. In terms of MSE, this number is much higher than any one of the 15-min or 30-min pedestrian flow simulation. Meanwhile, the training error margin expanded up to (-23.91, 26.62), and the error margin of the testing section also got increased slightly.

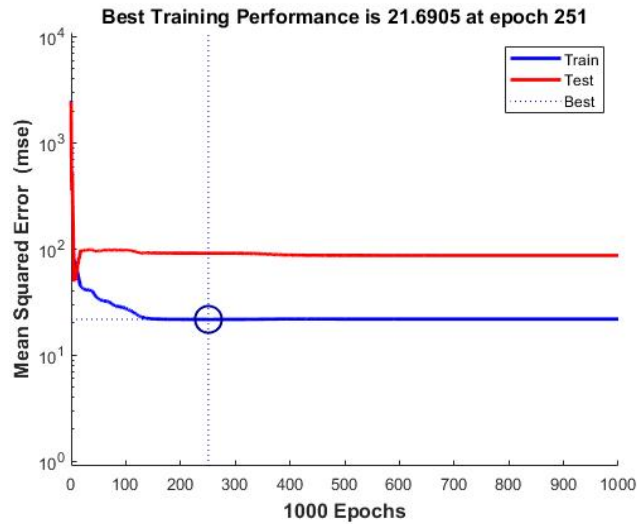


Figure 51

Training Performance of 60-Minute NOP Simulation

It found to be that the model performed well due to a training R of 0.9771, testing R of 0.9199, and an R of 0.9637 for the whole model. Correspondingly, as demonstrated in Figure 54, the time-series response also approximately fitted the real situation with few errors.

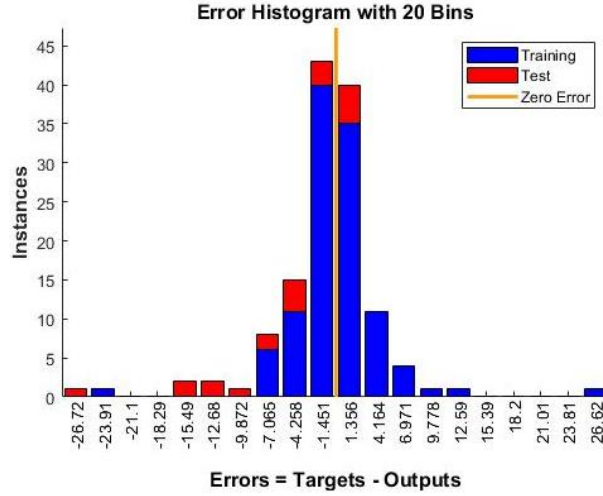


Figure 52

Error Histogram of 60-Minute NOP Simulation

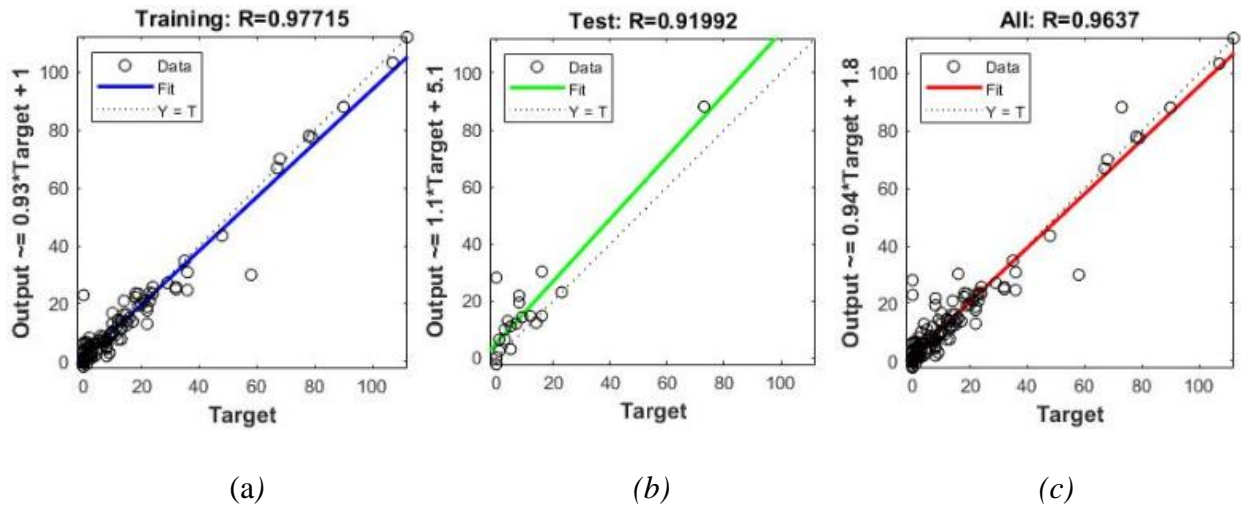


Figure 53 (a)-(c)

Regressions of 60-Minute NOP Simulation

As far as the 60-min vehicle flow simulation, when the best training performance reached at epoch 199 with a total of 521, as shown in Figure 55, the MSE of 97.77 was even less than the one associated with the 30-min vehicle flow simulation. However, the testing section showed a wider margin of error than training, according to Figure 56.

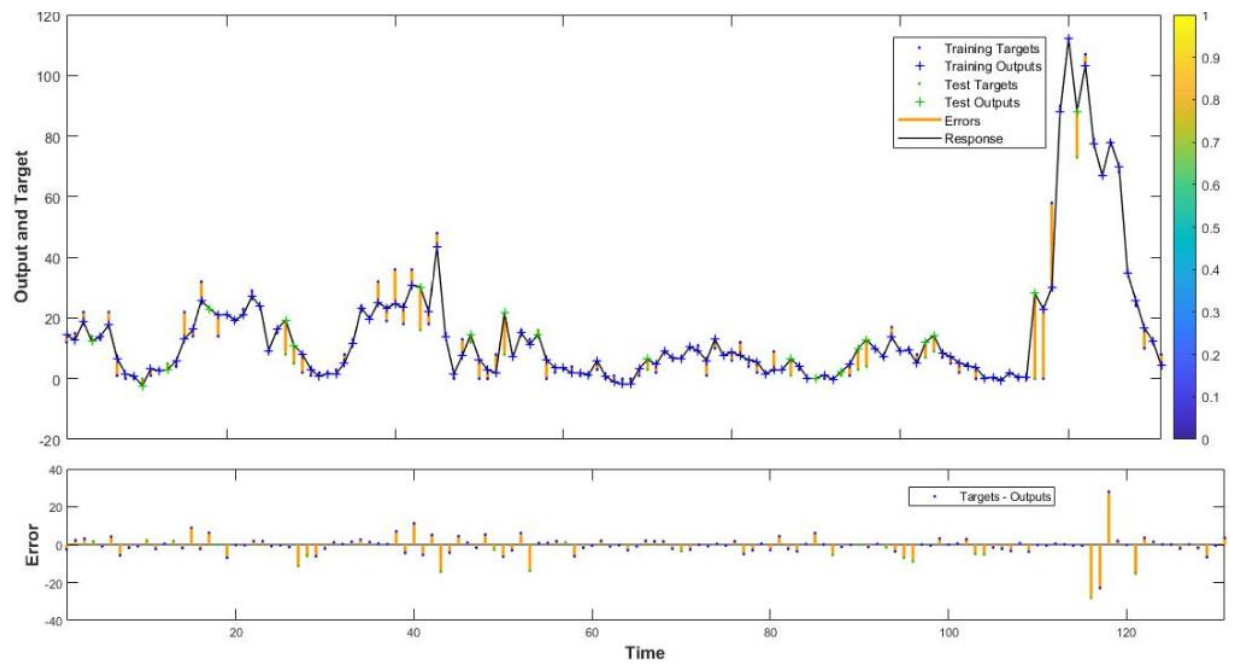


Figure 54

Time-Series Response of 60-Minute NOP Simulation



Figure 55

Training Performance of 60-Minute NOV Simulation

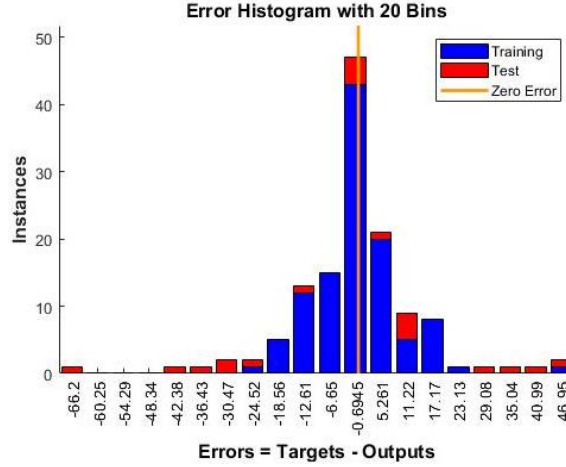


Figure 56

Error Histogram of 60-Minute NOV Simulation

In Figure 57 (a)-(c), it's indicated that 60-min vehicle flow simulation put up an excellent performance with a training R of 0.9813 but seemed to be a poor performance with the testing section due to a relatively low testing R of 0.8269. Overall, since the model R is 0.9584, the model still performed acceptably.

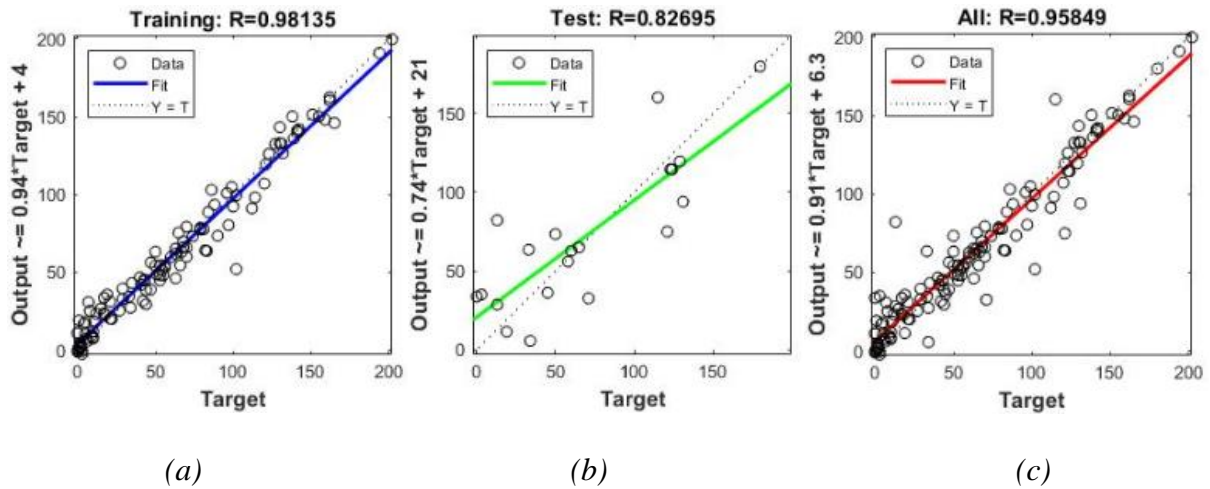


Figure 57 (a)-(c)

Regressions of 60-Minute NOV Simulation

As shown in Figure 58, the time-series response of the 60-min vehicle simulation can represent the real scenario of vehicle flow due to significant errors rarely occurred in the series.

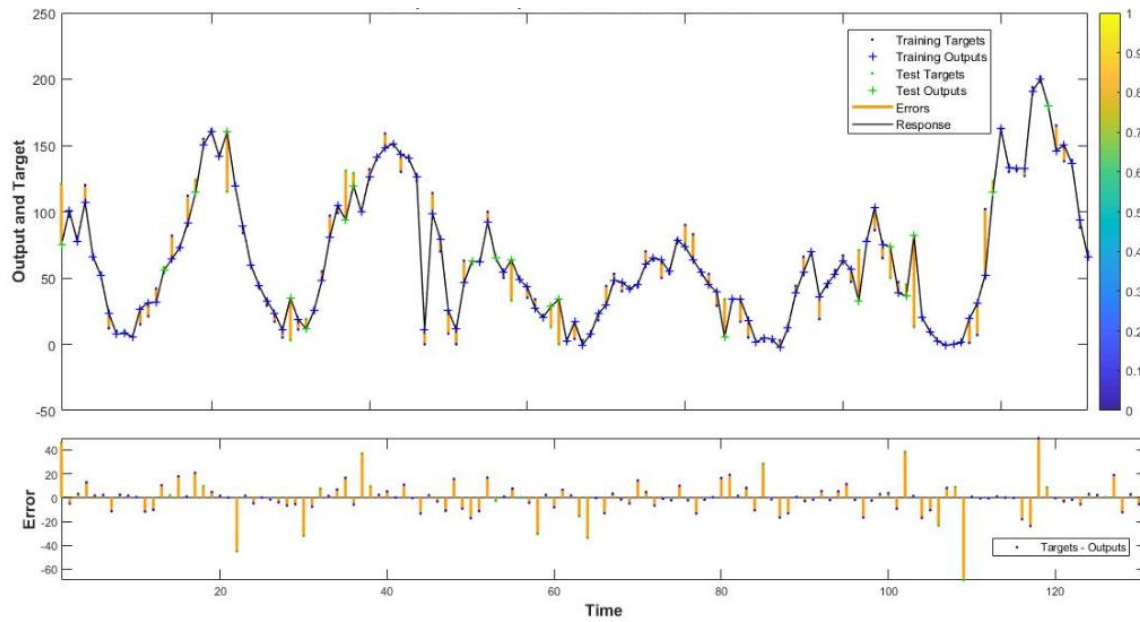


Figure 58

Time-Series Response of 60-Minute NOV Simulation

In summary, it's still hard to determine which pedestrian flow simulation model or vehicle flow simulation model performs the best merely relying on correlation coefficients and mean squared errors. To find out the most applicable model for simulating pedestrian and vehicle flow, respectively, in this study, a corresponding series of additional testing was conducted for model validation purposes.

Table 18

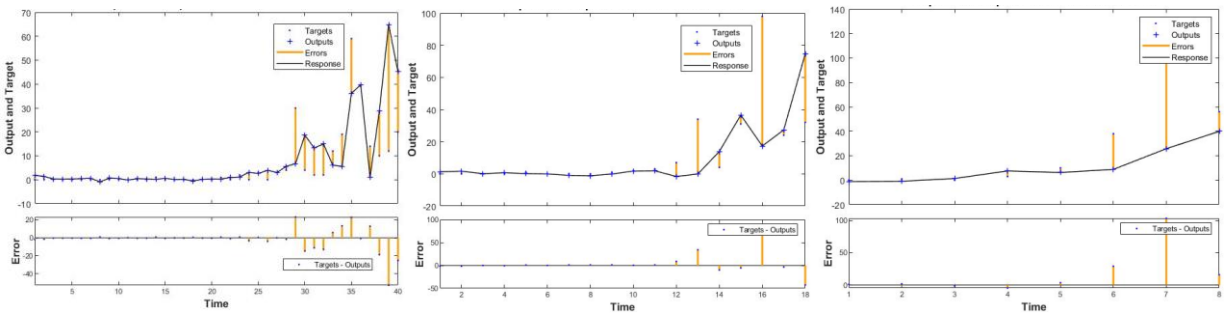
Performance Parameters of Model Simulations

Categorization	R Values of Regression			MSE	
	Training	Test	All	Training	Testing
15-Min NOP	0.9757	0.7659	0.9283	1.94	26.95
30-Min NOP	0.9862	0.9285	0.9747	3.3	23.39
60-Min NOP	0.9771	0.9199	0.9637	21.69	91.24
15-Min NOV	0.9428	0.8509	0.9268	22.68	62.07
30-Min NOV	0.9069	0.9335	0.9108	120.47	123.64
60-Min NOV	0.9813	0.8269	0.9585	97.77	844.33

Validation of ANN Traffic Simulation in MATLAB

As stated, the video data and meteorological data collected on March 19th of 2019 would use to validate the traffic flow simulation model in this study. Time-series responses of pedestrian and vehicle flow simulation for model validation and selection of the optimal traffic flow simulation model discussed in detail below.

Through Figures 59 (a)-(c) and Figure 60 (a)-(c), it's observed that all the simulated time-series response curves well fitted the actual curves in the previous stages but vary drastically in the second half step. However, overall, none of the fitting curves approximates the observed values as expected when conducting the additional testing of each model. In other words, the validation of the traffic flow simulation model didn't perform well. But, the relevant MSE and R associated with each additional testing take into the performance assessment of the corresponding model. For instance, using the data collected on March 19 of 2019 to validate all the traffic flow simulation models discussed above, two more columns of indicators, including MSE and R, were created for each model.



(a) 15-min NOP

(b) 30-min NOP

(c) 60-min NOP

Figure 59 (a)-(c)

Time-Series Response of NOP Simulation for Additional Testing

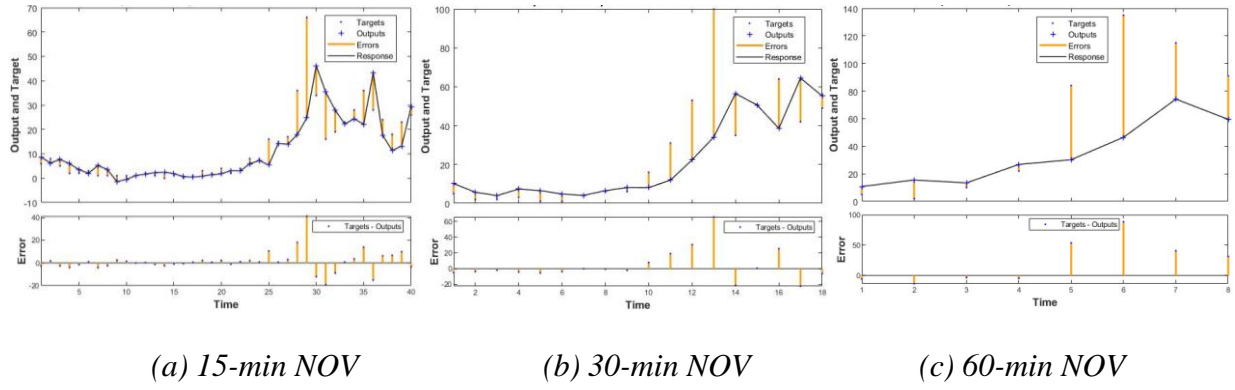


Figure 60 (a)-(c)

Time-Series Response of NOV Simulation for Additional Testing

As shown in Table 18, the original MSE for either training or testing purpose of each simulation varies in a wide range of (1.94, 844.33). In Table 19, a logarithmic transformation converts both the existed and newly created MSE values into corresponding numbers lying between 0.29 and 7.91.

Table 19

Performance Parameters with Additional Testing for All the Simulations

Categorization	R Values of Regression			LogMSE		Additional Testing		Sum_LogMSE
	R_Training	R_Testing	R_All	MSE_Training	MSE_Testing	LogMSE	R	
15-Min NOP	0.9757	0.7659	0.9283	0.29	1.43	2.08	0.5571	3.80
30-Min NOP	0.9862	0.9285	0.9747	0.52	1.37	2.73	0.4457	4.62
60-Min NOP	0.9771	0.9199	0.9637	1.34	1.96	3.17	0.7359	6.47
15-Min NOV	0.9428	0.8509	0.9268	1.36	1.79	1.94	0.7621	5.09
30-Min NOV	0.9069	0.9335	0.9108	2.08	2.09	2.62	0.6978	6.79
60-Min NOV	0.9813	0.8269	0.9585	1.99	2.93	2.99	0.8123	7.91

R-value or R^2 (Cameron et al., 1997; Asuero et al., 2006) is commonly used to represent the goodness of fit between predicted and observed values associated with a specific non-linear prediction model. In this study, it indicated that 30-min time-series BP neural network simulation model for pedestrian flow has the best goodness of fit with an R of 0.9747 if only taking the

model's goodness of fitting into consideration when running pedestrian flow simulation.

Similarly, the 60-min vehicle flow simulation model performs the best goodness of fitting with an R of 0.9585.

Previous studies (Vaisla et al., 2010; Mozo et al., 2018) indicated that mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE) widely applied for determining the performance of neural networks models. In this study, based on MSE theory, another transformation, logarithmic MSE (LogMSE), was brought into for choosing the optimal time-series BP traffic flow simulation model. In detail, a summation of LogMSEs retrieved from model training, testing, and additional testing was the final quantitative indicator of model performance. That is to say that the smaller the summation of LogMSEs is, the better performance the corresponding traffic flow simulation model has.

Under those studies mentioned above and the relevant results demonstrated in Table 19, Figures 61 and 62, both 15-min pedestrian and 15-min vehicle flow simulation models have the least summation of LogMSEs. Therefore, we can conclude that a 15-minute unit time interval is the most applicable one to conduct time-series BP neural network traffic flow simulation at the designated research spot in this study. Beyond that, since the series of simulations were run in MATLAB R2018b software, the associated coding script for 15-min traffic simulation models is attached in Appendix B at the end of this thesis.

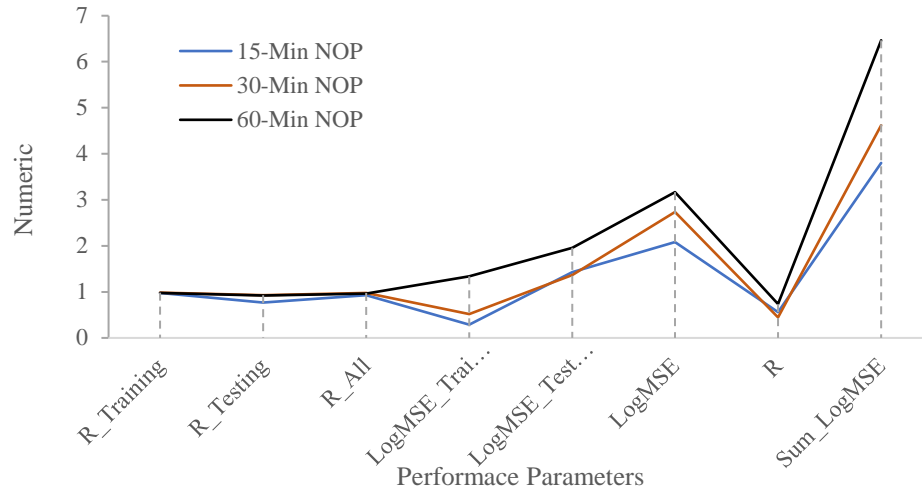


Figure 61

Performance Parameters of Pedestrian Flow Simulations

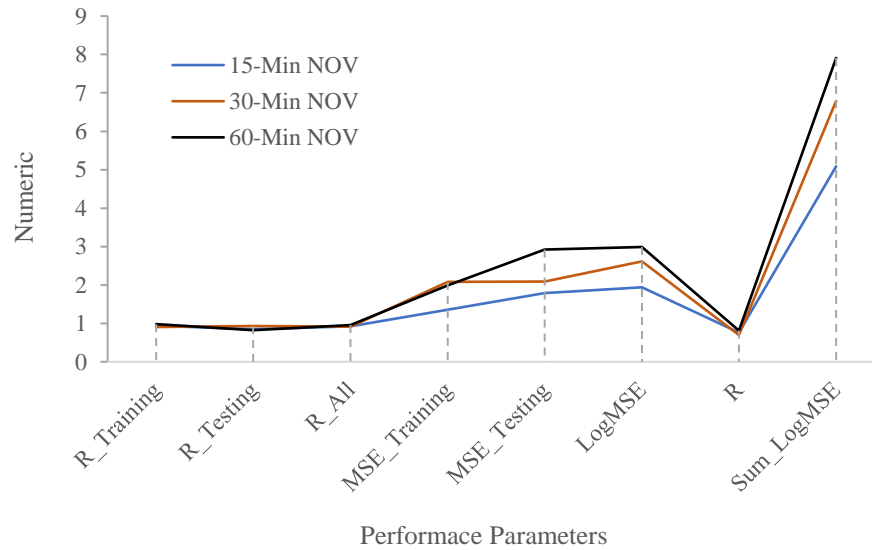


Figure 62

Performance Parameters of Vehicle Flow Simulations

Additionally, it's also found that the summation of LogMSE increases as the increment of the time interval. However, there is no such a default conclusion that a more excellent all R of the model refers to a better model performance due to no quantitative response relationship was found between model correlation coefficient R and summation of MSE. Without a doubt, the

optimal models for simulating pedestrian and vehicle flow determined in this study, but critical determinants out of all the variables described associated with each model haven't been found out yet. Therefore, for the sake of figuring out the vital determinants of the 15-min pedestrian/vehicle flow time-series BP neural network models, appropriate methods to conduct a sensitivity analysis of model input variables seems necessary.

Sensitivity Analysis of Model Input Variables

Theoretically, sensitivity analysis (SA) investigates how the variations of numerical model input factors or parameters contribute to the corresponding change with the model output. Following this broad definition, the type of approach, level of complexity, and purpose of SA vary significantly depending on the modeling domain and aims of the specific application (Pianosi et al., 2016). According to the relevant studies conducted by Pianosi et al. (2016), a variety of SA methods summarized and shown in Table 20 below.

Table 20

Methods of Sensitivity Analysis

SA Methods	Core Mathematical Expression	References
Perturbation and Derivatives	$s_i(\bar{x}) = \frac{\partial g}{\partial x_i} \bigg _{\bar{x}} c_i$	Devenish et al., 2012; Paton et al., 2013.
Multiple-starts Perturbation	$S_i = \frac{1}{r} \sum_{j=1}^r EE^j$	Morris, 1991; Rakovec et al., 2014.
Correlation and Regression Analysis	$S_i = correlation(x_i, y); S_i = b_i \frac{SD(x_i)}{SD(y)}$	Harper et al., 2011; Singh et al., 2014.
Reginal Sensitivity Analysis	$S_i = \max_{x_i} F_{x_i y_b}(x_i y_b \in Y) - F_{x_i y_{nb}}(x_i y_{nb} \in Y) $	Young et al., 1978; Tang et al., 2007b.
Variance-based Analysis	$S_i^F = \frac{V_{x_i}[E_{x_{-i}}(y x_i)]}{V(y)}$	Cukier et al., 1973; Norton, 2015.
Density-based Analysis	$S_i = \underset{x_i}{stat\ divergence}[f_y, f_{y x_i}(\cdot x_i)]$	Liu et al., 2006; Peeters et al., 2014

In this study, due to its simplicity, correlation analysis was applied to conduct sensitivity analysis between the input factors and output of the 15-min time-series BP neural network simulation model for traffic flow. To be specific, Pastres et al. (1999) proposed that the

Spearman rank correlation coefficient (SPCC) or partial rank correlation coefficient (PRCC) can use for nonlinear but monotonic relationships. Meanwhile, visual inspection is an informal though effective way to assess the sensitivity response of the input/output sample. Then, using Minitab, Spearman correlation analysis method was introduced to analyze the sensitivity between model input factors and output for the traffic flow simulation model. The details and relevant results associated with this analysis are shown in Table 21 and 22, respectively. In the third column of each table mentioned above, all correlation coefficient values converted from numbers saved in the second column to be positive by taking the absolute values. Apparently, in terms of the most influential model inputs of 15-min pedestrian flow simulation, with a descending order, they were the number of vehicles, humidity, temperature, atmosphere pressure, amount of rainfall or snow, vehicle waiting time, pedestrian waiting time, wind speed, train blocking time, and the number of trains, respectively. Similarly, with regards to the 15-min NOV simulation, its distribution of the most influential model inputs had almost the same tendency as 15-min NOP simulation, apart from two inputs, train blocking time and the number of passengers, exchanging the order.

Overall, it shows that the number of vehicles was the most sensitive parameter for 15-min time-series BP neural network pedestrian flow simulation. Meanwhile, the number of pedestrians crossing the selected intersection also responded sensitively to the model output of the 15-min vehicle flow simulation in this study. Furthermore, results of sensitivity analysis revealed that meteorological parameters were also the predominant input factors, exactly proving the assumption that weather factors may have an impact on the dynamic movements of pedestrians and vehicles. Specifically, the humidity was the most sensitive meteorological model input of both pedestrian and vehicle flow simulations.

Table 21*Spearman Correlation Coefficients between Inputs and Outputs of 15-min NOP Simulation*

Input Factors	NOP	Abs_NOP	Ranking
NOV	0.775	0.775	1
H (%) /100	-0.527	0.527	2
T(°F)	0.218	0.218	3
AP (in)	0.207	0.207	4
AOR/S (in)	-0.168	0.168	5
VWT (S)	0.131	0.131	6
PWT (S)	0.108	0.108	7
WS (mph)	-0.027	0.027	8
TBT (S)	0.005	0.005	9
NOT	0.003	0.003	10

Table 22*Spearman Correlation Coefficients between Inputs and Outputs of 15-min NOV Simulation*

Input Factors	NOV	Abs_NOV	Ranking
NOP	0.775	0.775	1
H (%) /100	-0.570	0.570	2
T(°F)	0.237	0.237	3
AP (in)	0.200	0.200	4
AOR/S (in)	-0.142	0.142	5
VWT (S)	0.076	0.076	6
PWT (S)	0.070	0.070	7
NOT	0.035	0.035	8
WS (mph)	-0.017	0.017	9
TBT (S)	0.015	0.015	10

Summary of Traffic Flow Simulation and Sensitivity Analysis

According to all the findings detected from both time-series BP neural network traffic flow simulations and sensitivity analysis conducted with the optimal models using Spearman correlation method, the results can summarize in the following:

1. It found that a 15-minute time interval is the most suitable one to conduct small-scale and dynamic neural network traffic flow simulations in this study. In other words, the 15-min

time-series BP neural network as a traffic flow simulation model is the most applicable one found in this study.

2. Regarding the goodness of fit of each simulation model, research results indicated that 30-min time-series BP neural network as a pedestrian flow simulation model had the best goodness of fit with an R of 0.9747. Similarly, the 60-min vehicle flow simulation model reached the best goodness of fitting with an R of 0.9585.
3. The sensitivity analysis executed on the optimal models revealed that model inputs, including the number of pedestrians and number of vehicles, respectively, responded most sensitively to outputs of vehicle flow and pedestrian flow simulations. In the meantime, with a descending order, weather parameters such as humidity, temperature, atmosphere pressure, and amount of rainfall or snow, as partial model inputs were also sensitive to the simulation model output.

Fuzzy Comprehensive Evaluation of Proposed Construction Schemes

Proposed Construction Schemes Associated with this Study

Based on the thorough analysis of video data and feedback of public surveys, it found that periodical standing cargo train indeed took both pedestrians and vehicle inconvenience to cross the intersection of South University Street and railway tracks in Normal, Illinois, USA. Except for the circumstance that there would always be a long wait for crossing availability when cargo trains blocked the way, pedestrians' dangerous behaviors, including passing through the train carriages or crossing the intersection when stop levers down, were also detected from the videos, as shown in Figure 63. Since uncontrolled situations such as cargo trains are getting longer and longer, the number of trains is also increasing. So it's tough to adjust the corresponding train schedules, some other thoughts urgently needed to propose for dealing with

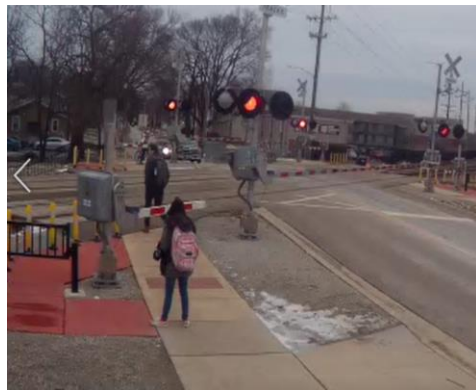
the severe safety hazards and providing both pedestrians and drivers convenience for access when crossing the intersection. As a response to those concerns, in this study, engineering schemes were come up with to tackle those negative consequences. In this study, three plans, including a pedestrian overpass, and the alternative options consisting of the pedestrian underpass and multi-use underpass proposed at the same time.



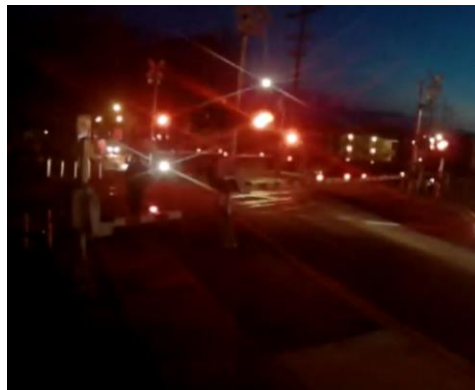
(a)



(b)



(c)



(d)

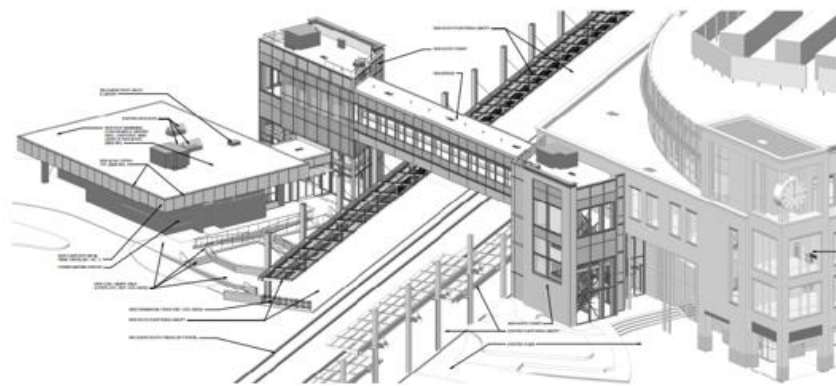
Figure 63 (a)-(d)

Negative Consequences of Periodical Standing Cargo Trains

In terms of the cases described associated with the Uptown Normal Pedestrian Railway Crossing Project, in total, there were six alternatives come up. And three of them show in Figure 64. Including enclosed pedestrian overpass, public overpass with a bridge, public overpass with

overlook/plaza, public underpass, public underpass with enhanced access, and open underpass with a park. Due to open access, several images associated with the study cited in this study.

Correspondingly, as shown in Figure 65, a matrix was created to compare different proposed schemes using semantic rating rules based on screening criteria like safety enhancement, economic development promotion, environmental concerns, cost, etc. Although the last plan listed in the figure determined as the optimal design for real construction, however, there was no information showing any quantitative analysis among criteria associated with all the six plans. In other words, a semantic-based quantitative comprehensive evaluation method can be taken into consideration when conducting a similar feasibility study of infrastructure projects.



(a) Enclosed Pedestrian Overpass



(b) Public Underpass with Plaza




(c) Open-air Pedestrian Overpass

Figure 64 (a)-(c)

Proposed Alternatives of Uptown Normal Project

Screening Criteria	No-Build	Enclosed Passenger Overpass Alt #2	Public Overpass with Bridge Alt #3a	Public Overpass with Overlook/Plaza Alt #3b	Public underpass Alt #4a	Public underpass with enhanced access Alt #4b	Public underpass with park Alt #4c
Enhances Safety for All	Poor	Poor	Good	Good	Good	Very Good	Very Good
Adverse travel for pedestrians	Poor	Poor	Good	Good	Very Good	Very Good	Very Good
Adverse travel for bicyclists	Poor	Poor	Fair	Fair	Very Good	Very Good	Very Good
Adverse travel for passengers	Very Poor	Very Good	Good	Good	Very Good	Very Good	Very Good
Connectivity/mobility for pedestrians	Poor	Poor	Good	Good	Good	Very Good	Very Good
Connectivity/mobility for bicyclists	Poor	Poor	Poor	Poor	Fair	Good	Very Good
Connectivity/mobility for passengers	Very Poor	Fair	Fair	Fair	Good	Very Good	Very Good
Public transportation access	Fair	Very Good	Good	Good	Good	Good	Good
Wayfinding requirements	Poor	Fair	Good	Good	Good	Very Good	Very Good
Promotes transportation opportunities for all	Poor	Fair	Good	Good	Good	Very Good	Very Good
Promotes economic development	Very Poor	Poor	Poor	Good	Fair	Good	Very Good
Environmental concerns	Very Good	Good	Good	Good	Fair	Fair	Fair
Aesthetic compatibility	Fair	Very Poor	Very Poor	Poor	Fair	Good	Very Good
Consistent with Master Plan	Very Poor	Very Poor	Fair	Good	Fair	Very Good	Very Good
Operational impacts during construction	Very Good	Good	Fair	Poor	Poor	Poor	Poor
Potential permanent right-of-way impacts	Very Good	Fair	Fair	Fair	Fair	Fair	Fair
Construction cost	\$	\$\$	\$\$\$	\$\$\$	\$\$\$	\$\$\$\$	\$\$\$\$
Operation & maintenance costs	\$	\$\$\$	\$\$\$	\$\$\$	\$\$\$	\$\$\$	\$\$\$
OVERALL RATING	Poor/Fair	Poor/Fair	Fair	Fair/Good	Fair/Good	Good	Good

 Recommended alternative

Normal Pedestrian Railroad Crossing Project

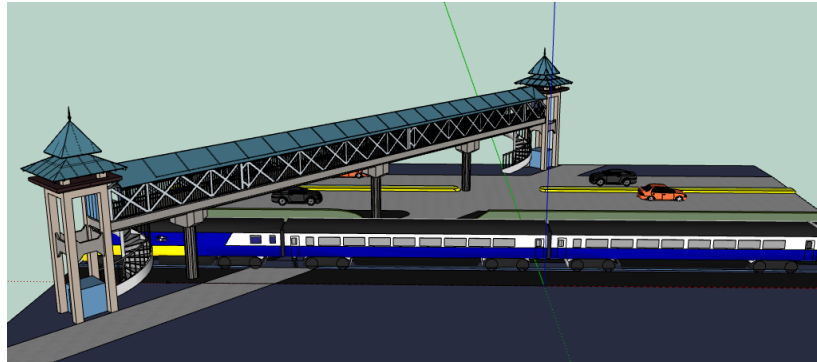
Figure 65

Screening Results and Recommendation of the Proposed Schemes for Uptown Normal Project

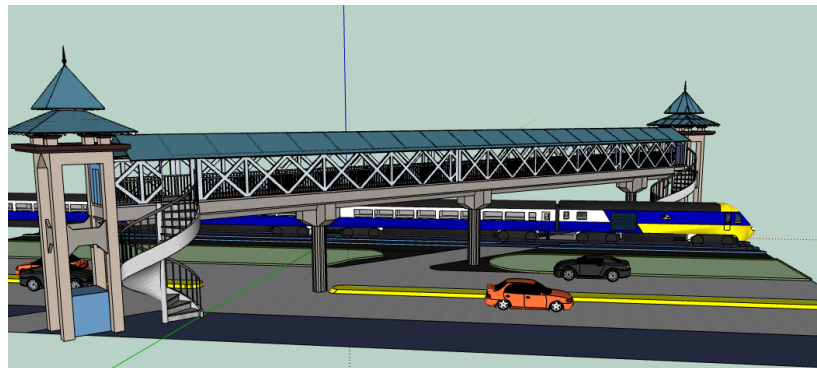
1. Pedestrian Overpass/Bridge

As demonstrated in Figure 66, a conceptual truss pedestrian bridge initially drawn in SketchUp 2019 based on the public feedback about which type of overpass they liked best in the online questionnaire. The actual situation shows that it doesn't make sense if the pedestrian overpass just built over the railways because of the short span of the bridge. Therefore, the adjacent street named Beaufort should also be taken into consideration when designing the railway-crossing facility. So, this is the reason why overpass has such a wide span and three piers supporting the load of the deck. Due to limited space near the bridgeheads, two spiral staircases creatively proposed to connect the floor and ground. In the meantime, two lifts also were designed to be respectively located at each end of the bridge to provide accessibility for the

disabled. Concerning the roofs of both bridge deck and elevator shafts, they were designed as a taper to eliminate the load of snow cover appearing in winter as much as possible.



(a)



(b)



(c)

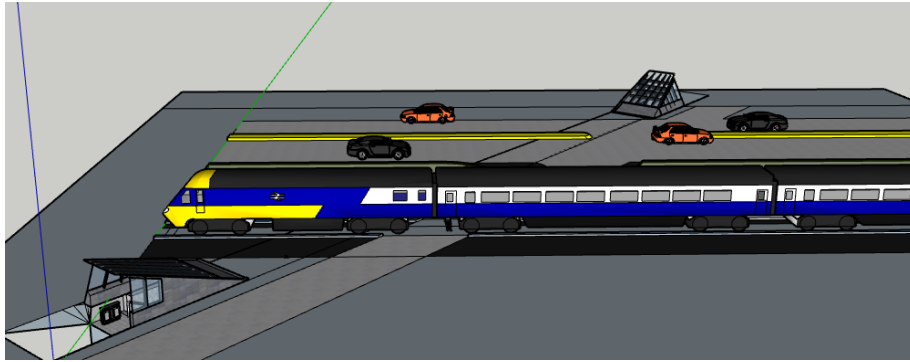


(d)

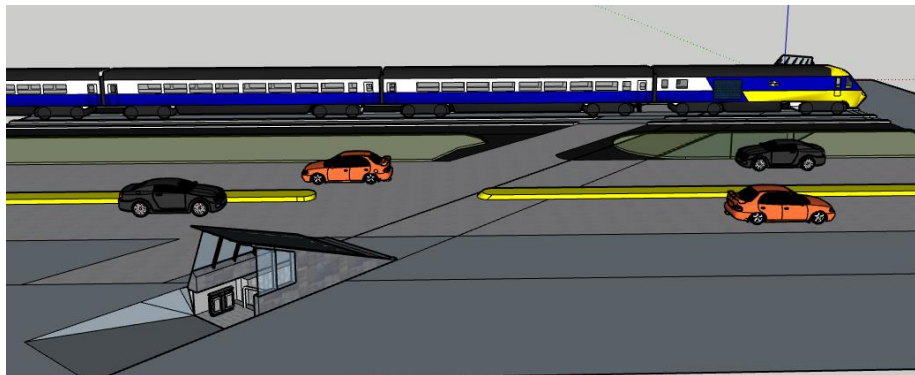
Figure 66 (a) – (d)

Scheme 1: Pedestrian Overpass

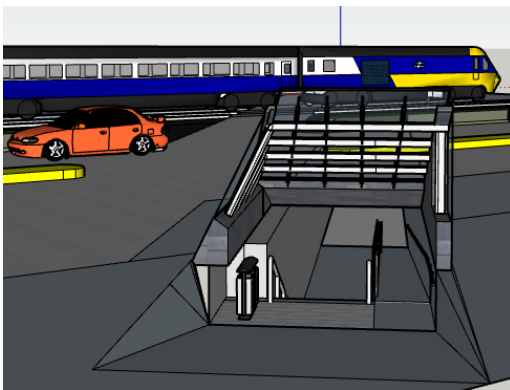
2. Pedestrian Underpass/Tunnel



(a)



(b)



(c)



(d)

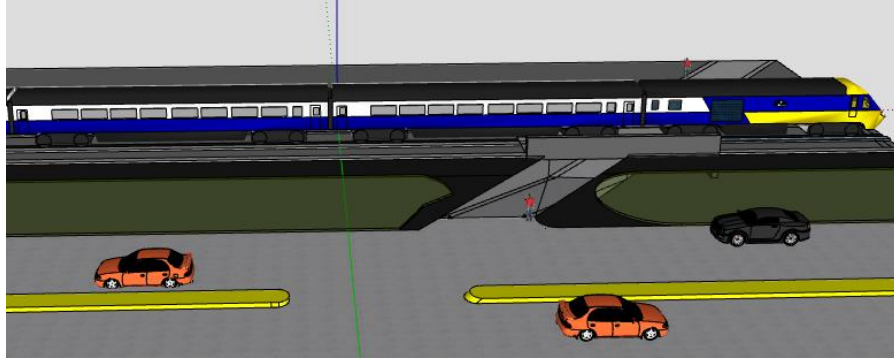
Figure 67 (a) - (d)

Scheme 2: Pedestrian Underpass

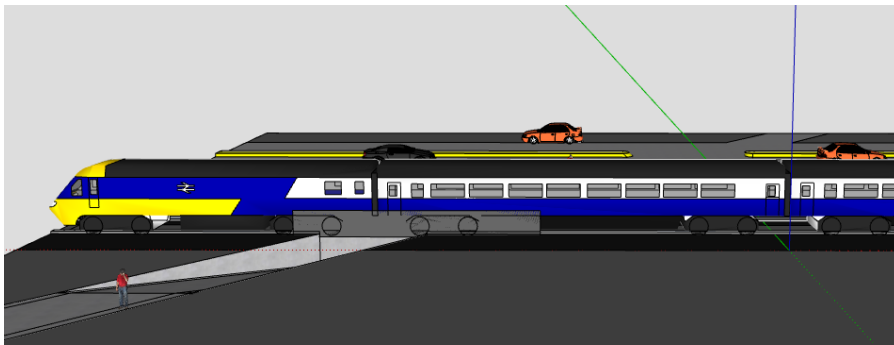
If only taking pedestrians and bicyclists into account when designing the underpass, due to the influence of traffic on Beaufort Street to them, a long pedestrian underpass has come up. Roughly, it's a 240 x 8 x 8-inch cuboid corridor with two tilting roofs on the inlet and outlet. So, pedestrians and bicyclists are free to cross these two intersections at any time. However, for vehicles passing across the intersection of rail tracks and S. University Street, they still must wait for access when the train arrives or stands on the tracks periodically. When the number of waiting vehicles increases too much, there will also be traffic congestion for Beaufort Street sometimes. Hence, a better design is needed to remove all the inconvenience caused by pedestrians, bicyclists, and drivers.

3. Multi-use Underpass/Tunnel

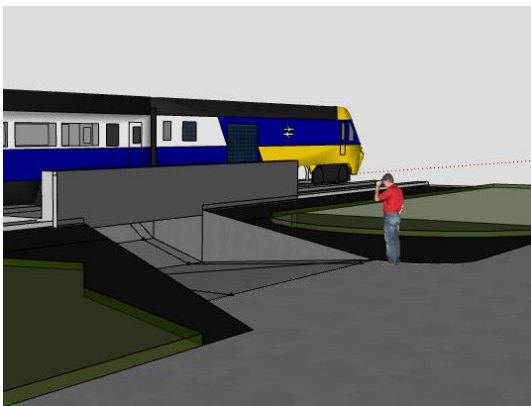
It happens that there is perhaps a similar case that one more underpass/tunnel design put forward. Still, this construction scheme can simultaneously provide easy access to pedestrians, bicyclists, and drivers who need to cross the intersection of railway tracks and S. University Street. This tunnel will construct beneath the rail tracks to reduce the construction difficulty and associated impacts on the traffic on the adjacent street and lessen the budget. This proposed multi-use underpass constitutes two slope sections and one cuboid corridor under the tracks. Also, the most challenging work with this underpass is that a flat railway bridge is necessarily needed to support the load of passing trains. By the means, it brings an extreme construction difficulty and a high cost for regular railway services due to a busy train schedule at the hypothetical construction site.



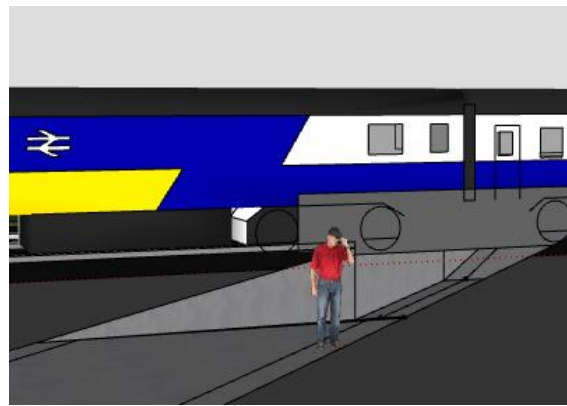
(a)



(b)



(c)



(d)

Figure 68 (a)-(d)

Scheme 3: Pedestrian and Vehicle Underpass

Of course, several critical factors, including social impacts, economic impacts, environmental impacts, budget, project schedule, etc. associated with each construction scheme

mentioned above will be deeply discussed in the following chapters when conducting the comprehensive evaluation of all these three plans. In this study, while the underlying mathematical principles of methods such as AHP and FEC have introduced in the methodology part, the following will continuously discuss the detailed implementation of these means in MATLAB R2018b for practical application of choosing the optimal construction scheme for decision-making.

Principles of AHP and FEC in MATLAB

1. Coding for Conducting AHP in MATLAB

Based on the mathematical principles introduced in the methodology of analytic hierarchy process in the study, except the first two steps that determination of hierarchical structure and pairwise comparison matrix K won't achieve in MATLAB, the other two phases, including calculation of priority vectors w_i and principal eigenvector value λ_{max} of matrix K , and conducting consistency test of matrix K , would be executed by the programming shown in Figure 69. The pairwise comparison matrix K can be determined easily according to the experts' experience. However, the consistency test has a strong capacity to avoid the subjectivity, since the proposed matrix K will be accepted only when the corresponding consistency ratio is 0.10 or less.

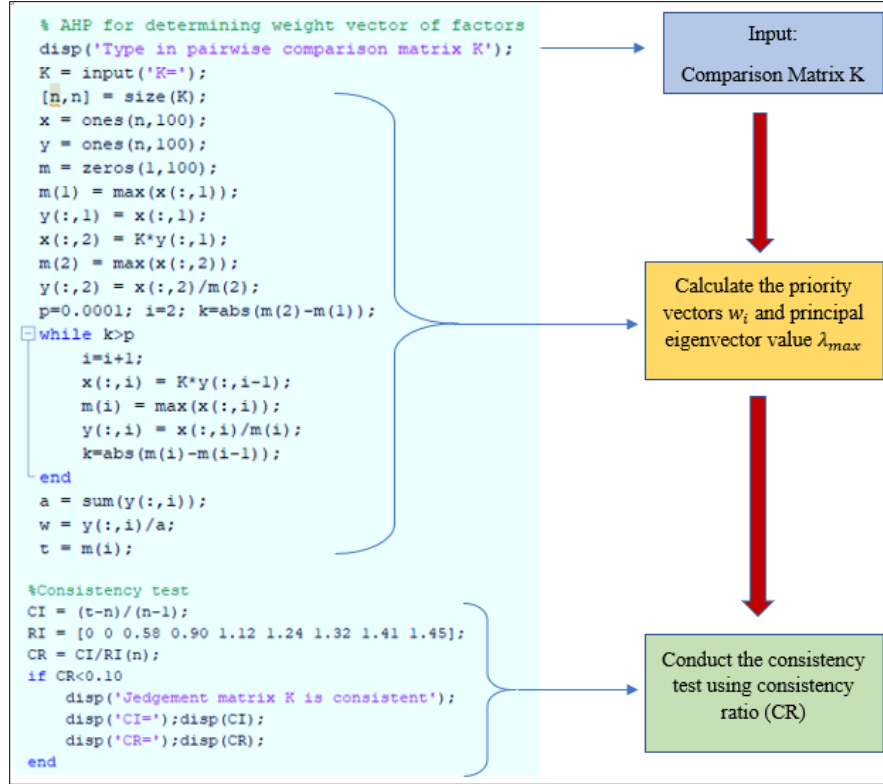


Figure 69

Principles of Conducting AHP in MATLAB

2. Coding for Conducting FEC in MATLAB

As demonstrated in Figure 70, four main steps of the fuzzy comprehensive evaluation of proposed construction schemes would be done in MATLAB using the matched code edited and compiled for this study. One precondition that the input arguments for this piece of code should be entirely quantitative. In other words, if there is a certain amount of qualitative or semantic judgment for criterion or factors of proposed alternatives, they must be converted to quantitative ones using appropriate linear scaling rule. Therefore, the precision of this transformation significantly affects the reliability of the final comprehensive evaluation of all the schemes. It's of vital importance in pointing out that this transformation can do by using statistical methods if permitted. For instance, given specific numerical rating scales to the corresponding semantic

judgments of involved criterion or factors, a questionnaire can be sent out for experts to rate the benchmark, which was judged in a qualitative way using the newly determined numerical scales.

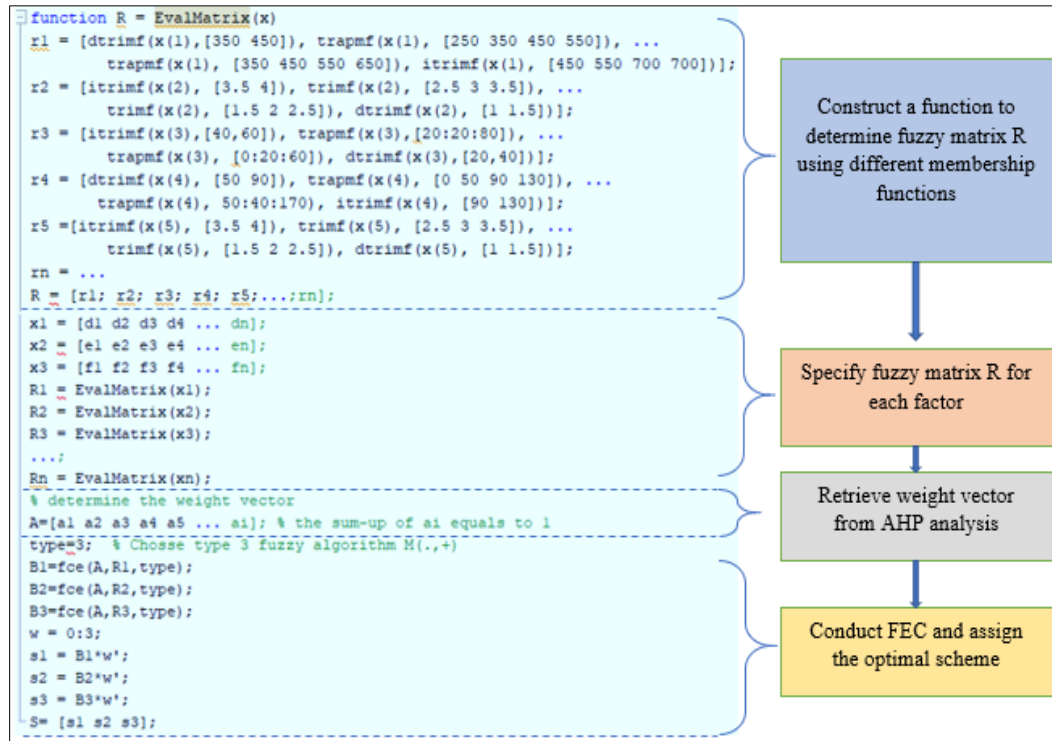


Figure 70

Principles of Conducting FEC in MATLAB

In terms of details of the code shown in this figure, looking down from the top, the first section is about making a customized MATLAB function containing a set of specified membership functions that would use for fuzzifying the quantitative judgments to fuzzy matrices. Details will be discussed in the next few paragraphs when conducting real applications associated with this study. Then, all the quantified judgment sets for all possible construction schemes will type in as the input arguments of the customized MATLAB function. In this way, fuzzy matrices for all the plans would be generated automatically for subsequent use. On top of this, the weight vector retrieved from the AHP part should load then. Consequently, evaluation

membership sets of all the schemes and overall grading set can ultimately calculate to direct how reliable decisions should make when choosing the optimal plan for future practical application.

Implementation of AHP-based FCE in MATLAB

In this study, based on the data which mined from each construction scheme through conducting individual analysis for all the eight factors involved and analyzing similar cases, AHP-based fuzzy comprehensive evaluation was implemented in MATLAB R2018b. In this way, it determines the optimal construction scheme for future decision-making when a real scenario happens. For the eight influential factors, including social impacts (indicated by safety improvement level, mobility for pedestrians and drivers), economic impacts, environmental impact, budget, maintenance, and schedule, they were listed and shown in Table 23. Overall, as far as the implementation procedures in this study, there were five main steps included, as shown in the following:

Step 1: Determine factor set X.

$$\mathbf{X} = (x_1, x_2, x_3, \dots, x_8) = (\text{SIL}, \text{MFD}, \text{MFP}, \text{EDPL}, \text{EIL}, \text{B}, \text{AOMC}, \text{S}).$$

Table 23

Influential Factors with Construction Schemes

Categorization	Influential Factors	Abbreviations	Elements of Factor Set X
Social Impacts	Safety Improvement Level	SIL	x_1
	Mobility for Drivers	MFD	x_2
	Mobility for Pedestrians	MFP	x_3
Economic Impact	Economic Development Promotion Level	EDPL	x_4
Environmental Impact	Environmental Impact Level	EIL	x_5
Budget	All Costs Related	B	x_6
Maintenance	Annual Operation & Maintenance Costs	AOMC	x_7
Schedule	Project Duration	S	x_8

Step 2: Determine Weight Vector w for the Set X. As stated, the analytic hierarchy process would introduce for determining the weights of factors. According to Saaty's numerical rating

rules shown in Table 1, the importance of pairwise factors or criterion was judged in turn and listed in Table 24.

Table 24

Results of Pairwise Judgement According to Saaty's Rule

Criterion	SIL	MFD	MFP	EDPL	EIL	B	AOMC	S
SIL	1	5	4	7	5	1	6	3
MFD	1/5	1	1/2	2	1/3	1/5	3	1/2
MFP	1/4	2	1	3	1	1	7	2
EDPL	1/7	1/2	1/3	1	1/2	1/2	3	1/3
EIL	1/5	3	1	2	1	1	2	2
B	1	5	1	2	1	1	3	2
AOMC	1/6	1/3	1/7	1/3	1/2	1/3	1	1/3
S	1/3	2	1/2	3	1/2	1/2	3	1

Then, the numbers in Table 24 were moved to an 8-by-8 pairwise comparison matrix K.

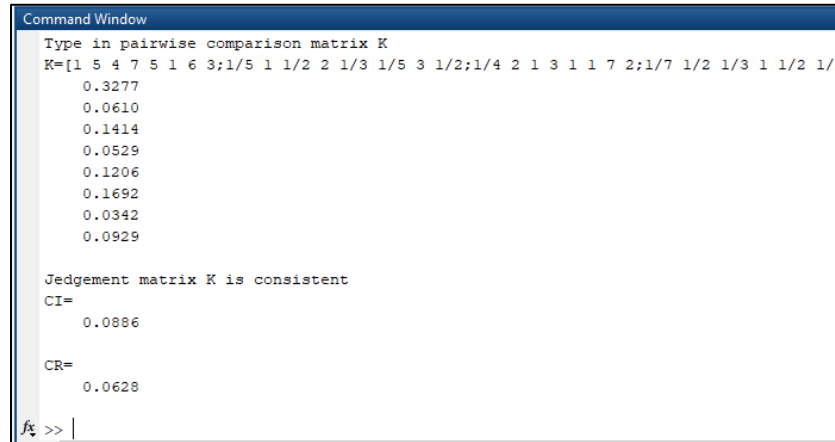
Importing matrix K into the MATLAB programming code shown in Figure 69, weight vector, and the result of consistency test were generated by MATLAB automatically, as demonstrated in Figure 71.

$$K = \begin{bmatrix} 1 & 5 & 4 & 7 & 5 & 1 & 6 & 3 \\ 1/5 & 1 & 1/2 & 2 & 1/3 & 1/5 & 3 & 1/2 \\ 1/4 & 2 & 1 & 3 & 1 & 1 & 7 & 2 \\ 1/7 & 1/2 & 1/3 & 1 & 1/2 & 1/2 & 3 & 1/3 \\ 1/5 & 3 & 1 & 2 & 1 & 1 & 2 & 2 \\ 1 & 5 & 1 & 2 & 1 & 1 & 3 & 2 \\ 1/6 & 1/3 & 1/7 & 1/3 & 1/2 & 1/3 & 1 & 1/3 \\ 1/3 & 2 & 1/2 & 3 & 1/2 & 1/2 & 3 & 1 \end{bmatrix}$$

In Figure 68, it found that consistency ratio (CR) was less than the limit value of 0.10, so the results of AHP for determining the weights of factors in this study were acceptable. At the meantime, weight vector $\mathbf{A} = (0.3277, 0.0610, 0.1414, 0.0529, 0.1206, 0.1692, 0.0342, 0.0929)$.

In Table 25, the weights of all the eight factors or criteria sorted. It indicated that the safety improvement level of the construction scheme has the highest weight value of 0.3277, which also proved that the original intention of these three engineering solutions was to mitigate the severe

safety concerns in the study area. Besides, budget, mobility for pedestrians, and environmental impact level took the second, third, and fourth place with a weight of 0.1692, 0.1414, and 0.1206, respectively.



```

Command Window
Type in pairwise comparison matrix K
K=[1 5 4 7 5 1 6 3;1/5 1 1/2 2 1/3 1/5 3 1/2;1/4 2 1 3 1 1 7 2;1/7 1/2 1/3 1 1/2 1/2
0.3277
0.0610
0.1414
0.0529
0.1206
0.1692
0.0342
0.0929

Judgement matrix K is consistent
CI=
0.0886

CR=
0.0628

fx >>

```

Figure 71
AHP Results Generated in MATLAB

Table 25
Weights of Factors and Corresponding Ranking

Factors	Elements in Factor Set	Weights	Ranking
SIL	x ₁	0.3277	1
MFD	x ₂	0.0610	6
MFP	x ₃	0.1414	3
EDPL	x ₄	0.0529	7
EIL	x ₅	0.1206	4
B	x ₆	0.1692	2
AOMC	x ₇	0.0342	8
S	x ₈	0.0929	5
Sum up		0.9999	

Step 3: Convert semantic judgments to quantitative ones. It's necessary to note that the precision of judgment transformations directly has an impact on the degree of evaluation subjectivity. As shown in Table 26, in the beginning, only three of the total eight factors quantitatively determined following appropriately professional analyses using techniques like

estimating of cost and duration in construction. The other five were merely rated qualitatively according to different rules of semantic scaling judgment empirically based on a few experts' reviews. Hence, it found that there was 62.5 % (5/8) of subjectivity involved, and low objectivity of 37.5% (3/8) if the final construction decision was merely made based on Table 26. In other words, this type of decision-making is deficient with reliability in the feasibility study of infrastructure projects. Therefore, one of the critical works of AHP-based FEC is to convert semantic judgments to quantitative for reducing the subjectivity of evaluation.

Table 26

Influential Factors and Corresponding Semantic and Quantitative Assessments to Construction Schemes

Influential Factors	Judgment Rating	Pedestrian Overpass	Pedestrian Underpass	Multi-use Underpass
SIL	(Very Weak, Weak, Moderate, Strong, Very Strong)	Strong	Strong	Very Strong
MFD	(Not at all, Low, Moderate, High, Extremely High)	Not at all	Not at all	Extremely High
MFP	(Not at all, Low, Moderate, High, Extremely High)	High	Extremely High	Extremely High
EDPL	(Low, Moderate, High)	High	Low	Moderate
EIL	(Low, Moderate, High)	Low	Moderate	High
B	Dollars	8.45 Million	11.5 Million	12.4 Million
AOMC	Dollars	35,000	15,000	10,000
S	Months	30	24	24

This study introduced the six-sigma methodology for determining the boundary values of each quantitative judgment interval to set up a reasonable quantitative judgment table for MOP and MOD. Using the simulated pedestrian and vehicle flow, it found that none of them distributed normally. Then, two data transformation methods, including squared root and Box-Cox (Box and Cox, 1964), introduced respectively. As for Box-Cox transformation, it can note as equation (33). The critical work is to estimate the coefficient lambda (λ) to reach a goal of transforming the data with non-normal distribution to normal distribution. Then, the six-sigma theory comes in

handy. However, in this study, neither 15-minute S-NOV nor 15-minute S-NOP distributes normally even after Box-Cox transformation, since the P-Values are still less than 0.005. Yet, the data distributions shown in Figures 72 and 74 have an approximately normal distribution.

Therefore, the six-sigma method could apply to this study.

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln y, & \lambda = 0 \end{cases} \quad (33)$$

To begin with, using MATLAB, lambdas λ_1 and λ_2 for transforming S-NOV and S-NOP were estimated as 0.3172 and -0.1955, respectively. It's worth noting that the variable y in Box-Cox formula should be greater than zero all the time. In other words, if the minimum element of a given vector with non-normal distribution is negative, a positive number that is greater than the absolute value of this negative can be added to all the items in this vector. In this way, all the elements in the given vector can convert into positive for the Box-Cox transformation. So, this is the reason why '3' and '9' added to every element in the vectors of 15-minute S-NOP and S-NOV vectors.

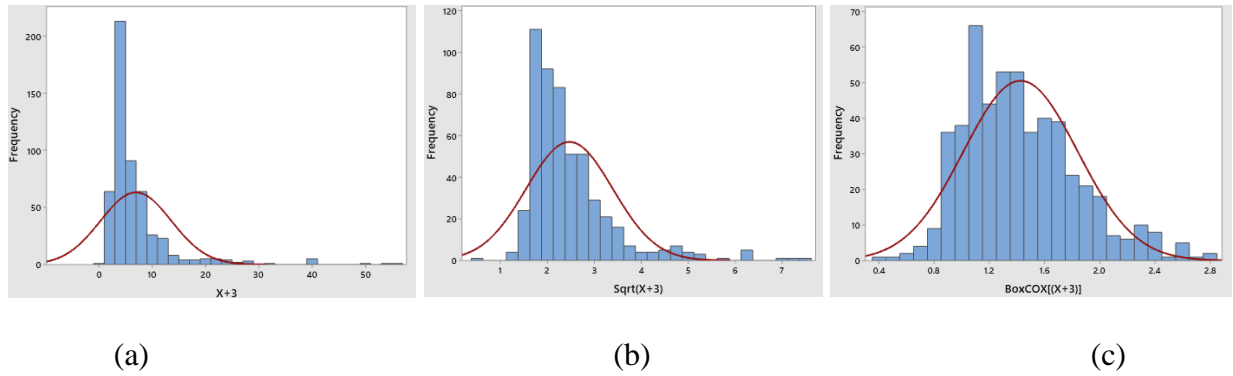


Figure 72(a)-(c)

Distribution of Different S-NOP Vectors

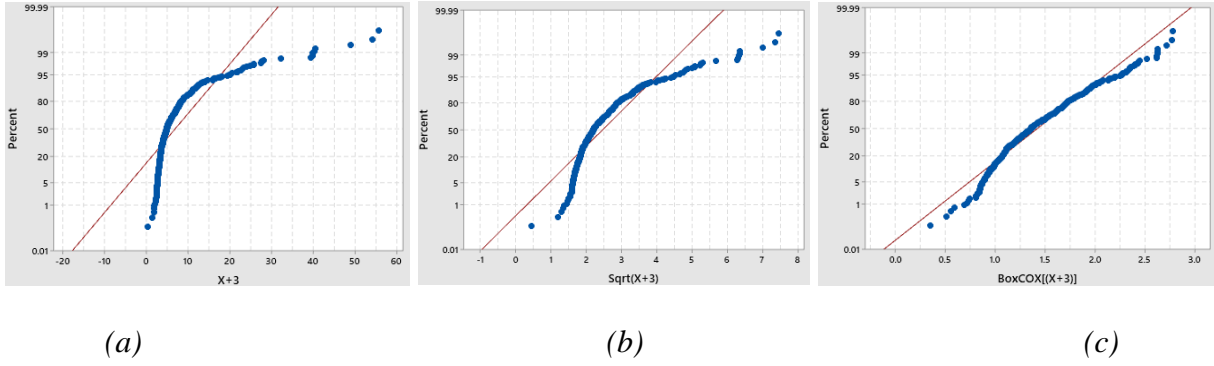


Figure 73
Probability Distribution of Different S-NOP Vectors

Table 27
Statistical Parameters of Different S-NOP Vectors

	Mean (μ_1)	St. Dev (σ_2)	P-Value
S-NOP + 3	6.94	6.64	< 0.005
Sqrt (S-NOP + 3)	2.45	0.92	< 0.005
Box-Cox (S-NOP + 3)	1.43	0.41	< 0.005

Table 28
Estimated Boundary Values for S-NOP Vectors

	Box-Cox	Converted (S-NOP+3)	Converted (S-NOP)
μ_1	1.43	5.33	2.33
$\mu_1 + \sigma_1$	1.84	9.80	6.80
$\mu_1 + 2\sigma_1$	2.26	19.61	16.61
$\mu_1 + 3\sigma_1$	2.67	43.73	40.73

Table 29
Statistical Parameters of Different S-NOV Vectors

	Mean (μ_2)	St. Dev (σ_2)	P-Value
S-NOV + 9	25.76	13.41	< 0.005
Sqrt (S-NOV + 9)	4.90	1.31	< 0.005
Box-Cox (S-NOV + 9)	5.43	1.45	< 0.005

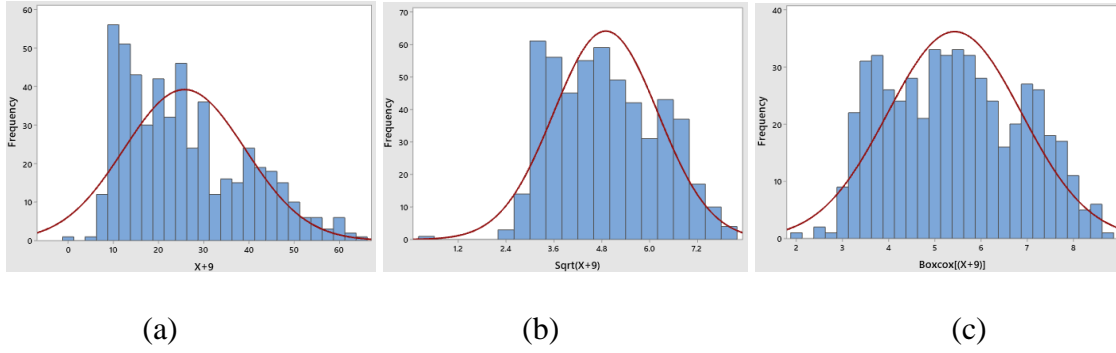


Figure 74 (a)-(c)

Distribution of Different S-NOV Vectors

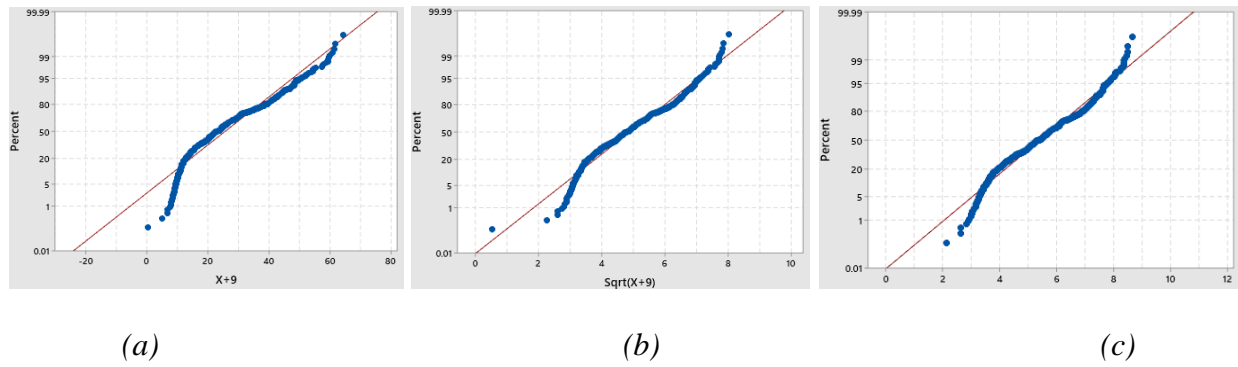


Figure 75 (a)-(c)

Probability Distribution of Different S-NOV Vectors

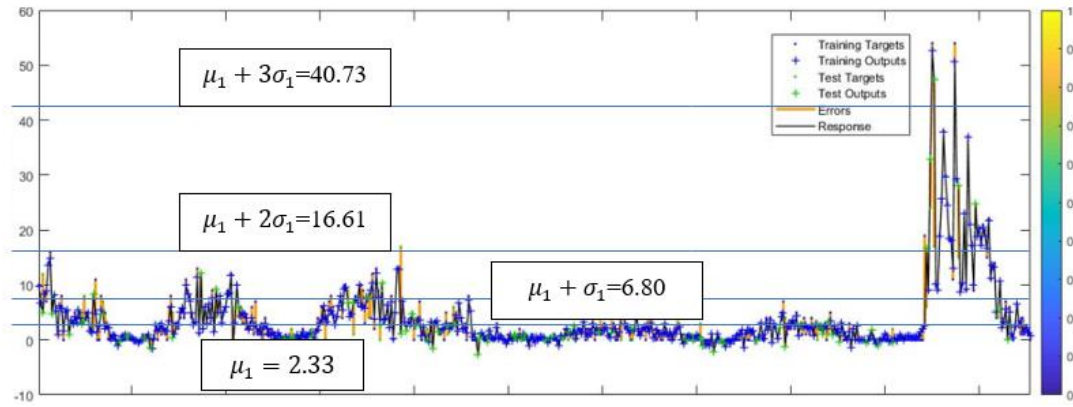
Table 30

Estimated Boundary Values for S-NOV Vectors

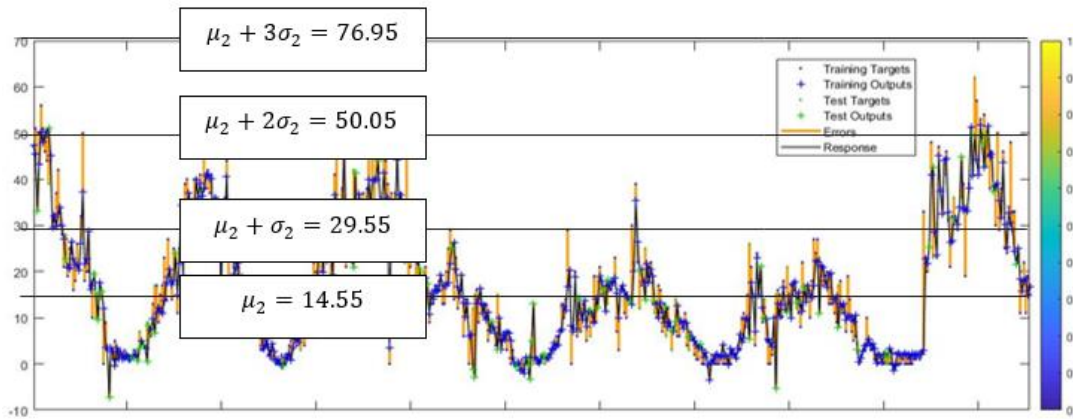
	Box-Cox	Converted (S-NOV+9)	Converted (S-NOV)
μ_2	5.43	23.55	14.55
$\mu_2 + \sigma_2$	6.88	38.55	29.55
$\mu_2 + 2\sigma_2$	8.33	59.05	50.05
$\mu_2 + 3\sigma_2$	9.78	85.95	76.95

According to all the results shown in Tables 27 – 30, the boundary values for MOP and MOV were determined appropriately. Then, accordingly, the evaluation scores for MOP and MOV were given, which demonstrate in Table 33. At the same time, based on the time-series plotting

of 15-minute S-NOP and S-NOV from ANN traffic flow simulations, Figure 76 (a)-(b) respectively demonstrate the visual division of boundaries for scoring using the six-sigma principles.



(a)



(b)

Figure 76 (a)-(b)

Quantitative Judgement Transmission for MFP and MPD Based on Traffic Simulation Results

In Table 31, a series of 10 typical case studies regarding feasibility studies of railway-crossing projects in several states of the U. S. conducted. Specifically, two critical criteria, including budget and construction duration, were collected. However, there were always some alternative schemes within each feasibility study listed in the table. In other words, the project budget varied

with the type of construction scheme as an underpass or overpass. Therefore, an average of all the budgets associated with proposed plans calculated to represent the estimated budget for each project. Then, the average estimated budget and the average construction duration were represented by μ_3 and μ_4 , respectively shown in Table 31 as well. Besides, the corresponding standard deviations σ_3 and σ_4 were also given for the estimated budget, and construction duration of each railway-crossing project. Then, applying the same principles used for determining the boundary values of MOP and MOV for scoring, the boundary values of estimated railway-crossing project budgets and construction durations were also identified, which demonstrate in Table 32. Afterward, in the quantitative judgment Table 33, a complete series of intervals for budget and construction duration was given for specific scoring for the proposed schemes in this study. With more concentration on the feasibility study of the Uptown Normal Railway Underpass Crossing Project, four different scoring intervals for AOMC defined accordingly. SIL, EDP, and EIL scored using an integer numerical scale from 1-4. In detail, the higher the SIL and EDP score, the better the corresponding schemes are. However, EIL takes the opposite scoring rule though it has the same scale.

Overall, evaluation set Y was also obtained, which is $Y = \{0, 1, 2, 3\}$. Specifically, '3' means the highest overall score a construction scheme, and '0' represents the lowest. For instance, based on the designed quantitative rating rules, as shown in Table 33, the most top overall rating will get when the proposed railway-crossing facility must reach a SIL and EDP of '4', a MOD of 76.95 vehicles per 15 minutes, a MOP of 40.73 pedestrians per 15-minute, the lowest EIL of '1', an estimated construction duration equal to or less than 15.32 million dollars, an estimated project budget equivalent to or less than 9.94 million dollars, and an AOMP less than 15,000 dollars simultaneously.

Table 31

Case Studies Regarding the Average Budget and Construction Duration of Railway-crossing Projects in the U.S.

Project Cases	Budget Underpass (Million)	Budget Overpass (Million)	Estimated Project Budget (Million)	Construction Schedule (month)
Route 39 Bridge in Aurora, MN	Not Given	5.7	5.7	24
Highway 47 Railway Separation in Anoka, MN	36.9	21.7	29.3	7
Railway Underpass Project, Woodhaven, Mich.	39	Not Given	39	Not Given
Brush College Rail Crossing Upgrades, Decatur, IL	13.5	Not Given	13.5	24
Uptown Normal Underpass Project, Normal, IL	24.6	Not Given	24.6	Not Given
Northgate Pedestrian Bridge Project, Seattle, WA	Not Given	16.5	16.5	Not Given
Highway 246 Bicycle/Pedestrian Bridge Project, Lompoc, CA	Not Given	7.5	7.5	Not Given
Belmont Bridge Replacement Project, Charlottesville, VI	27.3	15	21.15	24
Union City Boulevard / UPRR Grade Separation Project, Union City, CA	19.45	17.8	18.625	Not Given
McCord Road Railroad Underpass Project, Holland, Oregon	23.4	Not Given	23.4	16
Average	26.31	14.03	$\mu_3 = 19.93$	$\mu_4 = 19.00$
St. Dev.	9.10	6.20	$\sigma_3 = 9.99$	$\sigma_4 = 7.55$

Table 32

Estimated Boundary Values of Estimated Project Budgets and Construction Durations

Boundary Value	Budget (Million \$)	Boundary Value	Construction Dur. (Month)
μ_3	19.93	$\mu_4 - 0.5 \sigma_4$	15.23
$\mu_3 - 0.5 \sigma_3$	14.93	μ_4	19.00
$\mu_3 - \sigma_3$	9.94	$\mu_4 + 0.5 \sigma_4$	22.77

Table 33*Quantitative Judgement Rules of Construction Schemes*

Scores	Safety Improvement Level	Mobility for Drivers	Mobility for Pedestrians	Economic Development Promotion
3	4	≥ 76.95	≥ 40.73	4
2	3	[50.05, 76.95)	[16.61, 40.73)	3
1	2	[29.55, 50.55)	[6.80, 16.61)	2
0	1	< 29.55	< 6.80	1
Scores	Enviromental Impact Level	Constrution Dur. (MM)	Budget (Million \$)	Annual Operation & Maintainance Costs
3	1	≤ 15.23	≤ 9.94	$< 15,000$
2	2	(15.23, 19.00]	[9.94, 14.93)	[15,000, 30,000)
1	3	(19.00, 22.77]	[14.93 - 19.93)	[30,000, 45,000)
0	4	> 22.77	≥ 19.93	$\geq 45,000$

Using the proposed quantitative rating rule, as shown in Table 33, a few experts graded the proposed three construction schemes in this study, of which the results demonstrated in Table 34.

Table 34*Completely Quantitative Judgement of the Proposed Construction Schemes*

Influential Factors	Pedestrian Overpass	Pedestrian Underpass	Multi-use Underpass
Safety Improvement Level	2	2	3
Mobility for Drivers	0	0	3
Mobility for Pedestrians	2	3	3
Economic Development Promotion	3	1	2
Environmental Impact Level	1	3	4
Budget (Million \$)	8.45	11.5	13.15
Annual Operation & Maintenance Costs (\$)	35,000	15,000	10,000
Construction Duration (Month)	24	18	18

Step 4: Determine the fuzzy matrix using membership functions. Based on the factor set $X =$

$(x_1, x_2, x_3, \dots, x_8) = (SIL, MFD, MFP, EDPL, EIL, B, AOMC, S)$ and evaluation set $Y =$

$\{0, 1, 2, 3\}$, the first step in this phase is to determine the appropriate membership functions that

can use for the mapping of elements in factor set to evaluation set. As described, there are three

functions, including triangular, trapezoid, and Gaussian functions widely used. In this study, due to the long-winded description of the specific membership functions applied, if all of them were listed, only two discussed. To be precise, triangular membership function used to map element x_1 in the set X to evaluation set Y, as shown below:

(a) Triangular membership function $u_0(y_1)$ for calculating the membership of factor x_1 to element '0' in evaluations set.

$$u_0(x_1) = \begin{cases} 1, & x_1 \leq 1 \\ \frac{x_1-1.0}{1.5-1.0}, & 1 < x_1 < 1.5 \\ 0, & x_1 \geq 1.5 \end{cases} \quad (34)$$

(b) Triangular membership function $u_1(x_1)$ for calculating the membership of factor x_1 to element '1' in evaluations set Y.

$$u_1(x_1) = \begin{cases} 0, & x_1 < 1.5 \text{ or } x_1 > 2.5 \\ \frac{x_1-1.5}{2.0-1.5}, & 1.5 < x_1 < 2.0 \\ \frac{2.5-x_1}{2.0-1.5}, & 2.0 \leq x_1 < 2.5 \end{cases} \quad (35)$$

(c) Triangular membership function $u_2(x_1)$ for calculating the membership of factor x_1 to element '2' in evaluations set Y.

$$u_2(x_1) = \begin{cases} 0, & x_1 < 2.5 \text{ or } x_1 > 3.5 \\ \frac{x_1-2.5}{3.0-2.5}, & 2.5 < x_1 < 3.0 \\ \frac{3.5-x_1}{3.0-2.5}, & 3.0 \leq x_1 < 3.5 \end{cases} \quad (36)$$

(d) Triangular membership function $u_3(x_1)$ for calculating the membership of factor x_1 to element '3' in evaluations set Y.

$$u_3(x_1) = \begin{cases} 0, & x_1 < 3.5 \\ \frac{x_1-3.5}{4-3.5}, & 3.5 \leq x_1 < 4 \\ 1, & x_1 \geq 4 \end{cases} \quad (37)$$

However, when determining the membership of the element x_2 in the set X to items in Set Y, the other method, trapezoid membership function was applied. In detail, four derivative functions shown in the following:

(a) Trapezoid membership function $u_0(x_2)$ for calculating the membership of factor x_2 to element '0' in evaluations set Y.

$$u_0(x_2) = \begin{cases} 1, & x_2 < 29.55 \\ \frac{50.05-x_2}{50.05-29.55}, & 29.55 \leq x_2 < 50.05 \\ 0, & x_2 \geq 50.55 \end{cases} \quad (38)$$

(b) Trapezoid membership function $u_1(x_2)$ for calculating the membership of factor x_2 to element '1' in evaluations set Y.

$$u_1(x_2) = \begin{cases} \frac{x_2-14.55}{29.55-14.55}, & 14.55 \leq x_2 < 29.55 \\ 1, & 29.55 \leq x_2 < 50.55 \\ \frac{76.95-x_2}{76.95-50.55}, & 50.55 \leq x_2 < 76.95 \\ 0, & x_2 < 14.55 \text{ or } x_2 \geq 76.95 \end{cases} \quad (39)$$

(c) Trapezoid membership function $u_2(x_2)$ for calculating the membership of factor x_2 to element '2' in evaluations set Y.

$$u_2(x_2) = \begin{cases} \frac{x_2-29.55}{50.05-29.55}, & 29.55 \leq x_2 < 50.05 \\ 1, & 50.05 \leq x_2 < 76.95 \\ \frac{103.85-x_2}{103.85-76.95}, & 76.95 \leq x_2 < 103.85 \\ 0, & x_2 < 29.55 \text{ or } x_2 \geq 103.85 \end{cases} \quad (40)$$

(d) Trapezoid membership function $u_3(x_2)$ for calculating the membership of factor x_2 to element '3' in evaluations set Y.

$$u_3(x_2) = \begin{cases} 0, & x_2 \leq 50.05 \\ \frac{x_2-50.05}{76.95-50.05}, & 50.05 < x_2 < 76.95 \\ 1, & x_2 \geq 76.95 \end{cases} \quad (41)$$

As described above, due to four elements in the evaluation set Y there were also four membership functions needed when mapping each factor in set X to set Y. The membership functions of the rest six elements can be determined in the same manner since only were triangular and trapezoid functions utilized in this study.

Returning to the determination of the fuzzy set of each scheme based on the actual quantitative judgments, it significantly wastes the workforce if finding the corresponding fuzzy set of each plan using manual calculation one by one. In other words, if no advanced calculation methods proposed, there won't be such worthy of conducting this work since there would be a tremendous amount of workload. However, due to the compelling computerized calculation of MATLAB, the procedures of determining appropriate membership functions, and corresponding fuzzy set of each scheme can be executed in MATLAB R2018b when proper codes matched are edited and compiled. For instance, according to the principles of conducting FEC in MATLAB discussed above, a specific MATLAB function, shown in Figure 77, was written for calculating fuzzy sets for all the three proposed schemes. As clearly demonstrated in this figure, only two types of commands, including 'trimf' and 'trapmf' used when determining membership functions for constructing the needed fuzzy sets. Technically speaking, these two commands are precisely triangular and trapezoid membership functions, respectively, in MATLAB.

Step 5: Conduct a comprehensive evaluation. Based on quantitative judgments for each construction scheme and the MATLAB function for fuzzy set determination, a new piece of code, shown in Figure 78, was created to obtain the specific fuzzy sets and conduct the final evaluation of each scheme.

```

function R = FuzzySet(x)
r1 = [trimf(x(1),[0.5 1 1.5]), trimf(x(1), [1.5 2 2.5]), ...
      trimf(x(1), [2.5 3 3.5]), trimf(x(1), [3.5 4 4.5])];
r2 = [trapmf(x(2), [0 10 14.55 29.55]), trapmf(x(2), [10 14.55 29.55 50.05]), ...
      trapmf(x(2), [14.55 29.55 50.05 76.95]), trapmf(x(2), [29.55 50.05 76.95 103.85])];
r3 = [trapmf(x(3),[0 1 5 25]), trapmf(x(3),[1 5 25 55]), ...
      trapmf(x(3), [5 25 55 85]), trapmf(x(3),[25 55 85 115])];
r4 = [trimf(x(4),[0.5 1 1.5]), trimf(x(4), [1.5 2 2.5]), ...
      trimf(x(4), [2.5 3 3.5]), trimf(x(4), [3.5 4 4.5])];
r5 = [trimf(x(5), [3.5 4 4.5]), trimf(x(5), [2.5 3 3.5]), ...
      trimf(x(5), [1.5 2 2.5]), trimf(x(5), [0.5 1 1.5])];
r6 = [trapmf(x(6), [12 14 16 18]), trapmf(x(6), [10 12 14 16]), ...
      trapmf(x(6), [8 10 12 14]), trapmf(x(6), [6 8 10 12])];% Budget
r7 = [trapmf(x(7), [30000 45000 60000 75000]), trapmf(x(7), [15000 30000 45000 60000]), ...
      trapmf(x(7), [500 15000 30000 45000]), trapmf(x(7), [0 500 15000 30000])];%AMOC
r8 = [trapmf(x(8), [24 30 36 42]), trapmf(x(8), [18 24 30 36]), ...
      trapmf(x(8), [12 18 24 30]), trapmf(x(8), [6 12 18 24])];% Schedule
R = [r1; r2; r3; r4; r5; r6; r7; r8];

```

Figure 77

MATLAB Function for Determination of Fuzzy Sets

```

disp('Type in judgement set of scheme 1 as x1');
x1 = input('x1 =')
disp('Type in judgement set of scheme 2 as x2');
x2 = input('x2 =')
disp('Type in judgement set of scheme 3 as x3');
x3 = input('x3 =')
% The number of judgement sets strictly depends on how may the schemes
% or alternatives exactly are
R1 = FuzzySet(x1);
R2 = FuzzySet(x2);
R3 = FuzzySet(x3);
A=[0.3277, 0.0610, 0.1414, 0.0529, 0.1206, 0.1692, 0.0342, 0.0929];% Obtained from AHP
B1=A*R1;
B2=A*R2;
B3=A*R3;
w = 0:3;
s1 = B1*w';
s2 = B2*w';
s3 = B3*w';
R1
R2
R3
B1
B2
B3
S=[s1 s2 s3]

```

Figure 78

MATLAB Code for Comprehensive Evaluation of Proposed Construction Schemes

When running this MATLAB code, in the command window, the user would need to type in the required input arguments, which were the quantitative judgments of factors for each scheme

shown in Table 34. Consequently, fuzzy matrices for construction schemes, including pedestrian overpass, pedestrian underpass, and multi-use underpass, were automatically generated by MATLAB as matrix R_1 , R_2 , and R_3 , respectively. Since there were eight influential factors related to each construction scheme and four elements in evaluation set Y , there were eight rows and four columns in each fuzzy matrix R . Taking pedestrian overpass construction scheme as an example, the quantitative judgment of the first factor SIL, 2, was mapped to the evaluation set Y using triangular membership function in equation (34 - 37). Correspondingly, it fuzzified to a membership set $r_1 = (0 \ 1 \ 0 \ 0)$ comprising of four numbers, which only lie between 0 and 1. In other words, these four numbers were the memberships of SIL's quantitative judgment to the evaluation set $Y = \{3, 2, 1, 0\}$. In the same way, works were done with the adjacent factor MFD to get other membership set r_2 . One by one, finally, eight membership sets were obtained, which constituted the fuzzy matrix R_1 for evaluation of pedestrian overpass scheme after that, using the same idea, fuzzy matrices R_2 and R_3 were calculated in the end, which given in the following:

$$R_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0.1667 & 1 & 0.8333 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0.2250 & 1 \\ 0.3333 & 1 & 0.6667 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8333 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0.75 & 1 & 0.25 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0.4442 & 1 \\ 0 & 0 & 0.8333 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0.575 & 1 & 0.425 & 0 \\ 0 & 0 & 0.6552 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

By conducting the multiplication of the weight matrix A and the calculated fuzzy matrices R_1 , R_2 , and R_3 , respectively, three more evaluation membership sets, B_1 , B_2 , and B_3 , were created then. In detail, each element in set B represents the evaluation membership of each construction scheme to the elements including 0, 1, 2, and 3 in evaluation set Y in order. For instance, following the principle of maximum membership, 0.4784 is the maximum among the four elements in set B_1 , which indicates that the pedestrian overpass scheme should be more likely to score '1'. Similarly, pedestrian underpass and multi-use underpass schemes rated as '1' and '2' respectively, based on the given scoring scale 0-3 in the evaluation set Y in this study. Then, the final comprehensive assessment of these three proposed construction schemes can conclude that multi-use underpass was the most applicable. Both pedestrian overpass scheme and pedestrian underpass plans took second place.

$$B_1 = A \cdot R_1 = [0.0114 \ 0.4784 \ 0.3481 \ 0.4076]$$

$$B_2 = A \cdot R_2 = [0.0529 \ 0.5752 \ 0.4141 \ 0.3108]$$

$$B_3 = A \cdot R_3 = [0.2179 \ 0.2221 \ 0.6598 \ 0.3295]$$

However, according to the principle of maximum membership, each scheme was only scored based on the evaluation scores. Additionally, pedestrian overpass and pedestrian underpass schemes didn't differentiate due to the same score '1' graded. Hence, in search of the overall grade to each scheme, the paired product of b_i and y_i was summed up in order, as shown in equation (42).

$$s_k = \sum_{i=1}^4 b_i \cdot y_i \quad (42)$$

Then, an overall grade set S was acquired below:

$$S = [s_1 \ s_2 \ s_3] = [2.3974 \ 2.3359 \ 2.5303]$$

As a result, the overall grades for the pedestrian overpass, pedestrian underpass, and multi-use underpass schemes were 2.3974, 2.3359, and 2.5307. Ultimately, due to the multi-use scheme was graded the highest score, it found that the multi-use underpass construction scheme was the optimal one for real application if a railway-crossing facility will construct in the study area in the future. If this plan has no approval due to a higher budget, the pedestrian bridge scheme can be considered based on its overall grade of 2.3974 and the lowest budget.

Theoretically, the pedestrian underpass was not ideal for practical application.

Summary of Comprehensive Fuzzy Evaluation of Construction Schemes

In the light of appropriate methods included and evaluation results of the three proposed construction schemes associated with this study, some key points need to be underlined, which can be summarized below:

1. Due to the involvement of the analytic hierarchy process (AHP), the conventional method, experts' judgment, was not used to determine the weights of factors or criterion correspondingly related to the proposed construction schemes of the railway-crossing facility at the designated intersection. In this way, the AHP method significantly decreased the potential subjectivity contributing to the whole process of comprehensive evaluation when determining the pairwise comparison matrix K for calculation of fuzzy matrices because of its reliable consistency test. Specifically, the weights of those eight factors or criterion, including social impacts (indicated by safety improvement level, mobility for pedestrians and drivers), economic impact, environmental impact, budget,

maintenance, and schedule, were respectively calculated and stored in weight vector $A = (0.3277, 0.0610, 0.1414, 0.0529, 0.1206, 0.1692, 0.0342, 0.0929)$ in order. Due to the weight of safety impact level was the greatest with 0.3277, which indirectly proved the primary purpose of building the railway-crossing facility is to enhance the safety of both pedestrians and drivers who need to cross the selected intersection.

2. When qualitative or semantic judgments of factor sets exist, appropriate transformations should conduct to convert the unquantified judgments to quantitative ones using proper linear numerical rating scales. If permitted, statistical methods can introduce for practical application in this step, when questionnaire feedback with experts' quantitative judgments using the newly assigned numerical scales is accessible. Most importantly, the precision of converted judgments has a significant influence on the results of the comprehensive evaluation, because lower subjectivity involved means higher reliability the evaluation results will be. Yet, due to time urgency and the unavailability to get needed questionnaire results related to the works about transforming semantic judgments to quantitative, the researcher merely asked a few professional experts to help with this part. In other words, there might be few subjectivities involved because of the inadequate representativeness. Ultimately, an utterly quantitative judgment table obtained, as shown in Table 33.
3. Two types of membership functions were utilized in this study. Specifically, they were triangular and trapezoidal functions, respectively. Due to a great deal of work, if getting them with manual calculation, a customized function was written and loaded in MATLAB after all the specific parameters associated with corresponding membership functions determined. With some insights into those fuzzy matrices, the number of

columns should be exactly equivalent to the number of criteria or factors associated with the proposed schemes. Similarly, the quantity of rows in fuzzy matrices is supposed to be the same number of elements in the evaluation set Y.

4. If merely relying on the principle of maximum membership, occasionally, some proposed schemes may be graded the same score, then cause decision-maker difficulty of differentiating them, let alone sort them in order. Then, a derivative set S can be calculated based on the evaluation membership set B and evaluation set Y. In detail, the update can do by summing up the products of b_i and y_i , then saving them in new set S. By this means, the most considerable element in set S means the highest overall grade for the corresponding scheme, which indicates that it's the optimal scheme for practical application.
5. In this study, the results of the AHP-based fuzzy comprehensive evaluation model indicated that the multi-use underpass scheme is the most acceptable plan because of the highest overall evaluation score. In other words, this plan can implement to eliminate the severe safety hazards and decrease the significant negative impacts caused by periodically standing, and longer and longer cargo trains passing the study area.

CHAPTER VI: RESULTS AND CONCLUSIONS

In this study, a new technique named Fuzzy-Neural Comprehensive Evaluation Model (the FNCCEM) was proposed, built up, and applied successfully for the optimal selection of construction schemes in a feasibility study of a railway-crossing project in a college town. Concerning its principle, as mentioned already, this hybrid model consists of two basic machine learning methods, including neural networks and fuzzy logic. The primary purpose of this study is to investigate the applicability of the FNCCEM model for decision-making issues associated with the optimal selection of construction schemes in feasibility studies in infrastructure projects has reached. In the meantime, these two algorithms appropriately integrated some other critical techniques. To be specific, time-series theory and Bayesian theory combined with the BP neural networks for predicting the pedestrian and vehicle flow. Also, the analytic hierarchy process (AHP) combined with a fuzzy comprehensive evaluation of construction schemes proposed in the feasibility study of the railway-crossing facility project. Then, data collected and mined from traffic videos, online questionnaires, case studies, and online resources were cleaned and categorized accurately as model inputs. Ultimately, results obtained from time-series BP neural network traffic flow simulation, and AHP-based comprehensive evaluation for construction plans presented in the case discussed highlighted in the following:

1. It found that a 15-minute time span was the ideal one due to the quantitative model performance indicator, overall logarithmic mean squared error (Log-MSE), had the minimum values of 3.80 and 5.09 for pedestrian flow and vehicle flow simulation, respectively. In the meantime, a tendency captured that the wider the unit time is, the higher the overall Log-MSE of the corresponding traffic flow simulation model would be.

2. However, in terms of the goodness-of-fit of all six traffic flow simulation models proposed in this study, a 30-min pedestrian flow simulation model had the best degree of fitting with a correlation coefficient (R) of 0.9747 between the predicted values and observed values. Similarly, the 60-min vehicle flow simulation model performed the best due to a maximum R of 0.9585.
3. According to the result of sensitivity analysis of model inputs including pedestrian waiting time (PWT), vehicle waiting time (VWT), train blocking time (TBT), number of pedestrians (NOP), number of vehicles (NOV), number of trains (NOT), and homologous weather parameters consisting of temperature (T), wind speed (WS), humidity (H), atmospheric pressure (AP), amount of precipitation such as rainfall and snow (AOR/S), it indicated that NOP was significantly sensitive to the vehicle flow simulation, and NOV played the same role in pedestrian flow simulation. Additionally, it found that, with a descending order, weather parameters including humidity, temperature, atmosphere pressure, and amount of rainfall or snow, were more sensitive to the simulation model output than the other traffic variables except NOP and NOV.
4. Concerning the weights of criteria, results of AHP analysis manifested that safety improvement level was the most influential factor with the highest weight of 0.3277.
5. In this study, only two types of membership functions, triangular and trapezoidal functions, were utilized. MATLAB worked excellently to calculate the fuzzy membership matrix for each construction scheme.
6. Under the principle of maximum membership, given an evaluation set $Y = \{0 \ 1 \ 2 \ 3\}$, three proposed plans, including pedestrian overpass, pedestrian underpass, and multi-use

underpass, scored '1', '1', '2', respectively. In other words, if so, the third scheme should be optimally selected for practical application in the future.

7. However, due to the pedestrian bridge and tunnel schemes were scored the same, based on the calculated evaluation membership sets. Overall grading was conducted then for final decision-making, of which the result indicated that multi-use underpass was still the ideal construction design. Even pedestrian overpass and pedestrian underpass took the second and third places, respectively. In order, the overall scores were 2.3974, 2.3359, and 2.5307 for the pedestrian overpass, pedestrian underpass, and multi-use underpass, respectively.

In the end, several crucial conclusions can make through the exploration, build-up, and application of the FNCEM technique in this study. As is demonstrated below:

1. The artificial neural network (ANN) indeed is a powerful nonlinear prediction tool for small-scale traffic flow simulation. By integrating ANN and other techniques such as time-series theory and Bayesian regularization, three types of ANN models using 15-minute, 30-minute, and 60-minute timespan, were respectively developed for implementing small-scale simulations of pedestrian flow and vehicle flow. Then, the simulated pedestrian flows and vehicle flows were practically used as reliable requirements of mobility that a proposed railway-crossing facility should meet.
2. Analytical Hierarchy Processing (AHP), as an independent technique for decision-making problems, showed significant feasibility of being integrated with fuzzy logic for the selection of construction schemes in the feasibility study of infrastructure projects. In this way, the subjectivity in the determination of weights of evaluation criterion or

influencing factors to an infrastructure project is drastically avoided due to the required consistency test of AHP.

3. In sequential order, the outputs of the optimal ANN traffic simulation module were used as partial evaluation criterion or influencing factors of the infrastructure project in the AHP-based fuzzy evaluation module. Besides, considering other evaluation criteria/influencing elements, the ANN module and AHP-based fuzzy evaluation module jointly constitute the FNCEM model for optimal selection of construction schemes in the feasibility study of infrastructure projects.
4. In the literature, apart from the application of ANN in traffic simulation and the use of AHP-based fuzzy logic in the problems of optimal selection of construction schemes, no similar work has found the establishment of the FNCEM model for the optimal selection of construction scheme in feasibility study of infrastructure projects, especially for railway-crossing projects. A conclusion can be drawn that the proposed FNCEM model significantly has the potentials to enrich the methodology for the feasibility study of infrastructure projects.

However, there were also several limitations affiliated with this study, which are listed as follows:

1. In the phase of transforming semantic judgments to quantitative for developing the quantitative judgment matrix, one of the nonquantifiable critical influencing factors, political background, was not considered.
2. Experts' judgments were used to grade the proposed railway-crossing facilities using the quantitative judgment matrix developed in this study. However, a more scientific method, questionnaire, and statistical analysis were not applied to do so.

3. At present, no appropriate database has found accessibility for the validation of the FNCEM model.
4. In theory, the FNCEM model is tightly bonded with ANN and AHP-based fuzzy logic. However, due to plenty of steps involved to develop this model, if a Simulink program in MATLAB can be made for future application of this model, it will have higher applicability.

REFERENCES

- Abou-Zeid, Azza & Bushraa, Ashraf & Ezzat, Maged. (2007). Overview of Feasibility Study Procedures for Public Construction Projects in Arab Countries. *Journal of King Abdulaziz University-Engineering Sciences*, 18, 19-34.
- Ahmed, M. S. and Cook, A. R. (1979). Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques. *Transportation Research Board*, 722, 1–9.
- Asuero, A. G., Sayago, A., & González, A. G. (2006). The Correlation Coefficient: An Overview. *Critical Reviews in Analytical Chemistry*, 36(1), 41–59.
- B. N. Passow, D. Elizondo, F. Chiclana, S. Witheridge, and E. Goodyer (2013). Adapting Traffic Simulation for Traffic Management: A Neural Network Approach. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), the Hague, 1402-1407.
- Bi, H. K., Mao, T. L., Wang, Z. Q., and Deng, Z. G. (2016). A Data-driven Model for Lane-changing in Traffic Simulation. *Eurographics/ ACM SIGGRAPH Symposium on Computer Animation*.
- Box, G. E. P., and Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211-243.
- Brzozowska, A., Bubel, D., & Kalinichenko, A. (2019). Analysis of the Road Traffic Management System in the Neural Network Development Perspective. *Eastern-European Journal of Enterprise Technologies*, 98(3), 16–24.
- Buscema, M., Ferilli, G., & Sacco, P. L. (2017). What Kind of ‘World Order’? An Artificial Neural Networks Approach to Intensive Data Mining. *Technological Forecasting & Social Change*, 117, 46–56.

- Cameron, A. C., & Windmeijer, F. A. G. (1997). An R-Squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models. *Journal of Econometrics*, 77(2), 329–342.
- Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., Schaibly, J.H. (1973). Study of The Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients. I Theory. *J. Chem. Phys.* 59 (8), 3873-3878.
- Danish Farooq, & Janos Juhasz. (2019). Simulation-Based Analysis of the Effect of Significant Traffic Parameters on Lane Changing for Driving Logic “Cautious” on a Freeway. *Sustainability*, 21, 5976.
- Demšar, J., & Zupan, B. (2013). Orange: Data Mining Fruitful and Fun - A Historical Perspective. *Informatica* (03505596), 37(1), 55–60.
- Devenish, B., Francis, P.N., Johnson, B.T., Sparks, R.S.J., Thomson, D.J. (2012). Sensitivity Analysis of Dispersion Modeling of Volcanic Ash from Eyjafjallajökull in May 2010. *J. Geophys. Res.* 117.
- Ding, S. F., Jia, W. K., Su, C. Y., Liu, X. L., and Chen, J. R. (2010). An Improved BP Neural Network Algorithm Based on Factor Analysis. *Journal of Convergence Information Technology*, 5(4), 103-108.
- Elhalwagy, A. M., Ghoneem, M. Y. M. & Elhadidi, M. (2017). Feasibility Study for Using Piezoelectric Energy Harvesting Floor in Buildings’ Interior Spaces. *Energy Procedia*, 115, 114–126.
- Elkind, L. D. C., & Landini, G. (2018). *The Philosophy of Logical Atomism: A Centenary Reappraisal*. Palgrave Macmillan. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&db=cat00180a&AN=milner.2044698&site=eds-live&scope=site>

- Faye S., Chaudet C. (2016). Characterizing the Topology of an Urban Wireless Sensor Network for Road Traffic Management. *IEEE Transactions on Vehicular Technology*, 65(7), 5720-5725.
- Feng, C. Y., Wang, H. Y., Lu, N. J., Chen, T., He, H., Lu, Y., and Tu, X. M. (2014). Log-Transformation and Its Implications for Data Analysis. *Shanghai Archives of Psychiatry*, 26(2), 105–109.
- Feng, S., and Xu, L. D. (1999). Decision Support for Fuzzy Comprehensive Evaluation of Urban Development. *Fuzzy Sets and Systems*, 105(1), 1-12.
- Franzese, M., & Iuliano, A. (2019). Correlation Analysis. *Encyclopedia of Bioinformatics and Computational Biology*, 706–721.
- Frechtling, J., Silverstein, G., Zhang, X. D., et al. (2010). Feasibility Study Report for the National Center for Research Resources.
- French, J. (2016). The Time Traveler's CAPM. *Investment Analysts Journal*. 46 (2): 81–96.
- Fu, R., Zhang, Z., and Li, L. (2017) Using Istm and Gru Neural Network Methods for Traffic Flow Prediction. *Youth Academic Conference of the Chinese Association of Automation*. 324–328.
- Gaines, B. R. (1976). Foundations of Fuzzy Reasoning. *International Journal of Man-Machine Studies*, 8(6), 623-668.
- Gamut, L. T. F. (1991). *Logic, Language, and Meaning, Volume 1: Introduction to Logic*. University of Chicago Press. Pp. 156–157.
- Georgakellos, D. A. & Marcis, A. M. (2009). Application of the Semantic Learning Approach in the Feasibility Studies Preparation Training Process. *Information Systems Management* 26, 231-240.

- Gunduz, M., & Mohammad, K. O. (2020). Assessment of Change Order Impact Factors on Construction Project Performance Using Analytic Hierarchy Process (AHP). *Technological & Economic Development of Economy*, 26(1), 71–85.
- Hamed, M. M., Al-Masaeid, H. R., and Said, Z. M. B. (1995). Short-term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering*, 121(3), 249–254.
- Harper, E., Stella, J.C., Fremier, A. (2011). Global Sensitivity Analysis for Complex Ecological Models: A Case Study of Riparian Cottonwood Population Dynamics. *Ecol. Appl.* 21 (4), 1225-1240.
- Hendiani, S., & Bagherpour, M. (2019). Developing an Integrated Index to Assess Social Sustainability in Construction Industry Using Fuzzy Logic. *Journal of Cleaner Production*, 230, 647–662.
- Hu, G. X. (2005). Application of Fuzzy Optimization Theory in Bridge Scheme Selection. *Journal of Highway and Transportation Research and Development*, 22(7), 90-92.
- Huang, W. H., Song, G. J., Hong, H. k., and Xie, K. Q. (2014). Deep Architecture for Traffic Flow Prediction: Deep Belief Networks with Multitask Learning. *IEEE Transactions on Intelligent Transportation Systems* 15(5), 2191–2201.
- Hyari, Khaled & Kandil, Amr. (2009). The Validity of Feasibility Studies for Infrastructure Construction Projects. *Jordan Journal of Civil Engineering*, 3, 66-77.
- Ji, H. Y., Xing, Z. Y., Qin, Y. Jia, L. M., and Wu, G. J. (2011). Evaluation on Railway Passenger Service Quality Using Comprehensive Fuzzy Evaluation Model and Neural Network. *Journal of Central South University (Science and Technology)*, 42(1), 72-78.

- Ji, H.-Y & Xing, Z.-Y & Qin, Yong & Jia, Li-min & Wu, G.-J. (2011). Evaluation on Railway Passenger Service Quality Using Comprehensive Fuzzy Evaluation Model and Neural Network. *Journal of Central South University (Science and Technology)*, 42, 72-78.
- Jin, J., Li, M., & Jin, L. (2015). Data Normalization to Accelerate Training for Linear Neural Net to Predict Tropical Cyclone Tracks. *Mathematical Problems in Engineering*, 2015, 1–8.
- Justis, R. T. & Kreigsmann, B. (1979). The Feasibility Study as a Tool for Venture Analysis. *Business Journal of Small Business Management*, 17, 35-42.
- Konatowski, S., & Gologowski, M. (2017). Koncepcja Systemu Monitorowania Ruchu Pojazdów Drogowych. (Polish). *Przegląd Elektrotechniczny*, 93(10), 64.
- Krejcie, R. R. & Morgan, D. W. (1970). Determining Sample Size for Research Activities. *Educational and Psycho.*
- Labib, S. M. (2019). Investigation of the Likelihood of Green Infrastructure (GI) Enhancement along Linear Waterways or on Derelict Sites (DS) Using Machine Learning. *Environmental Modelling and Software*, 118, 146–165.
- Larsen, J. (1999). Introduction to Artificial Neural Network (1st edition). The Technical University of Denmark.
- Li, Y., Jiang, X., Zhu, H., He, X., Peeta, S., Zheng, T., & Li, Y. (n.d.). Multiple Measures-based Chaotic Time Series for Traffic Flow Prediction based on Bayesian Theory. *Nonlinear Dynamics*, 85(1), 179–194.
- Libiao Bai, Yi Li, Qiang Du, & Yadan Xu. (2017). A Fuzzy Comprehensive Evaluation Model for Sustainability Risk Evaluation of PPP Projects. *Sustainability*, (10), 1890.
- Liu, H., Sudjianto, A., Chen, W. (2006). Relative Entropy-Based Method for Probabilistic Sensitivity Analysis in Engineering Design. *J. Mech. Des.* 128, 326-336.

- Lu, H., Sun, Z., Qu, W., & Wang, L. (2015). Real-Time Corrected Traffic Correlation Model for Traffic Flow Forecasting. *Mathematical Problems in Engineering*, 2015, 1–7.
- McLean County Regional Planning Commission. (2012). Main Street Transportation Improvement Feasibility Study for Bloomington-Normal, Illinois. Retrieved from <https://www.normal.org/DocumentCenter/View/1252/Feasibility-Study-FINAL-DRAFT-to-Council-41612?bidId=>
- Ministry of Transportation & Infrastructure Customer Satisfaction Survey (2019). <https://engage.gov.bc.ca/govtogetherbc/consultation/ministry-of-transportation-and-infrastructure-customer-satisfaction-survey>
- Moore, D. S., Notz, W. I., & Flinger, M. A. (2013). *The Basic Practice of Statistics* (6th ed.). New York, NY: W. H. Freeman and Company.
- Morris, M. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33 (2), 161-174.
- Mozo, A., Ordozgoiti, B., & Gómez-Canaval, S. (2018). Forecasting Short-Term Data Center Network Traffic Load With Convolutional Neural Networks. *PLoS ONE*, 13(2), 1–31.
- Mukaka, M. M. (2012). Statistics Corner: A Guide to Appropriate Use of Correlation Coefficient in Medical Research. *Malawi Medical Journal*, 24(3), 69–71.
- Nabian, M. A., Meidani, H. (2017). Deep Learning for Accelerated Reliability Analysis of Infrastructure Networks. *Computer-Aided Civil and Infrastructure Engineering*. 33 (6): 443–458.
- Norton, J. (2015). An Introduction to Sensitivity Assessment of Simulation Models. *Environ. Model. Softw.* 69, 166-174.

- Okutani, I. and Stephanedes, Y. J. (1984). Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory. *Transportation Research Part B Methodological*, 18 (1), 1–11.
- Pastres, R., Chan, K., Solidoro, C., Dejak, C. (1999). Global Sensitivity Analysis of a Shallow-Water 3D Eutrophication Model. *Comput. Phys. Commun.* 117, 62-74.
- Paton, F.L., Maier, H.R., Dandy, G.C. (2013). Relative Magnitudes of Sources of Uncertainty in Assessing Climate Change Impacts on Water Supply Security for the Southern Adelaide Water Supply System. *Water Resour. Res.* 49 (3), 1643-1667.
- Peeters, L., Podger, G., Smith, T., Pickett, T., Bark, R., Cuddy, S. (2014). Robust Global Sensitivity Analysis of a River Management Model to Assess Nonlinear and Interaction Effects. *Hydrol. Earth Syst. Sci.* 18 (9), 3777-3785.
- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity Analysis of Environmental Models: A Systematic Review with Practical Workflow. *Environmental Modelling & Software*, 79, 214–232.
- Polson, N., and Sokolov V. O. (2017). Deep Learning for Short-term Traffic Flow Prediction. *Transportation Research Part C: Emerging Technologies*, 79, 1–17.
- Rakovec, O., Hill, M.C., Clark, M.P., Weerts, A.H., Teuling, A.J., Uijlenhoet, R. (2014). Distributed Evaluation of Local Sensitivity Analysis (DELSA), with Application to Hydrologic Models. *Water Resour. Res.* 50 (1), 409-426.
- Ristoski, P., Bizer, C., & Paulheim, H. (2015). Mining the Web of Linked Data with RapidMiner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 35(3), 142–151.
- Saaty, R. W. (1987). The Analytic Hierarchy Process – What It Is and How It Is Used. *Math Modeling*, 9(3-5), 161-176.

- Saaty, T., and Vargas, L. (2001). *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. Kluwer Academic Publishers.
- Sarigiannis, D. A. (2015). *Data Infrastructure and Data Mining Model of Internal Exposome*. Nature: Report – R.
- Schnell, T., Mohror, J. S., & Aktan, F. (2002). *Evaluation of Traffic Flow Analysis Tools Applied to Work Zones Based on Flow Data Collected in the Field*. Transportation Research Record, No.1811, 57.
- Simon, H. (2001). *Neural networks: A Comprehensive Foundation (Second Edition)*. Pearson Education, Inc. (Book)
- Singh, R., Wagener, T., Crane, R., Mann, M.E., Ning, L. (2014). *A Vulnerability Driven Approach to Identify Adverse Climate and Land Use Change Combinations for Critical Hydrologic Indicator Thresholds: Application to A Watershed in Pennsylvania, USA*. Water Resour. Res. 50, 3409-3427.
- Sitton, J. D., Zeinali, Y., & Story, B. A. (2019). *Design and Field Implementation of an Impact Detection System Using Committees of Neural Networks*. Expert Systems with Applications, 120, 185–196.
- Slimani, N., Slimani, I., Sbiti, N., & Amghar, M. (2019). *Traffic Forecasting in Morocco Using Artificial Neural Networks*. Procedia Computer Science, 151, 471–476.
- Sola, J., and Sevilla, Joaquin (1997). *Importance of Input Data Normalization for the Application of Neural Networks to Complex Industrial Problems*. IEEE Transactions on Nuclear Science 44(3):1464 – 1468.
- Southland Association. (2012). *Southland Drive Green Infrastructure Feasibility Study*.

- Sun, S. L., Zhang, C. S, and Yu, G. Q. (2006). A Bayesian Network Approach to Traffic Flow Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 7(1), 124–132.
- Sun, S., Zhang, C., Yu, G., Lu, N., & Xiao, F. (2004). Bayesian Network Methods for Traffic Flow Forecasting with Incomplete Data. *Lecture Notes in Computer Science*, 419.
- Swanson, D. C. (2017). *Signal Processing for Intelligent Sensor Systems with MATLAB* (2nd edition). Boca Raton, Florida: CRC Press.
- Tang, Y., Reed, P., Wagener, T., van Werkhoven, K. (2007b). Comparing Sensitivity Analysis Methods to Advance Lumped Watershed Model Identification and Evaluation. *Hydrol. Earth Syst. Sci.* 11, 793-817.
- Tao Zhou, & Yulin Zhou. (2015). Fuzzy Comprehensive Evaluation of Urban Regeneration Decision-making Based on Entropy Weight Method: A Case Study of Yuzhong Peninsula, China. *Journal of Intelligent & Fuzzy Systems*, 29(6), 2661–2668.
- The Town of Normal. (2019). The Underpass Project at Uptown Station: BUILD FY 201p Application. Retrieved from https://www.wglt.org/sites/wglt/files/201911/normal_build_fy2019_narrative_-_07-15-2019.pdf
- Ullah, I., Baharom, M. N. R., Ahmad, H., Wahid, F., Luqman, H. M., Zainal, Z., & Das, B. (2019). Smart Lightning Detection System for Smart-City Infrastructure Using Artificial Neural Network. *Wireless Personal Communications*, 106(4), 1743–1766.
- Vaisla, K. & Bhatt, A. (2010). An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting. *International Journal on Computer Science and Engineering*, 2(6), 2104-2109.

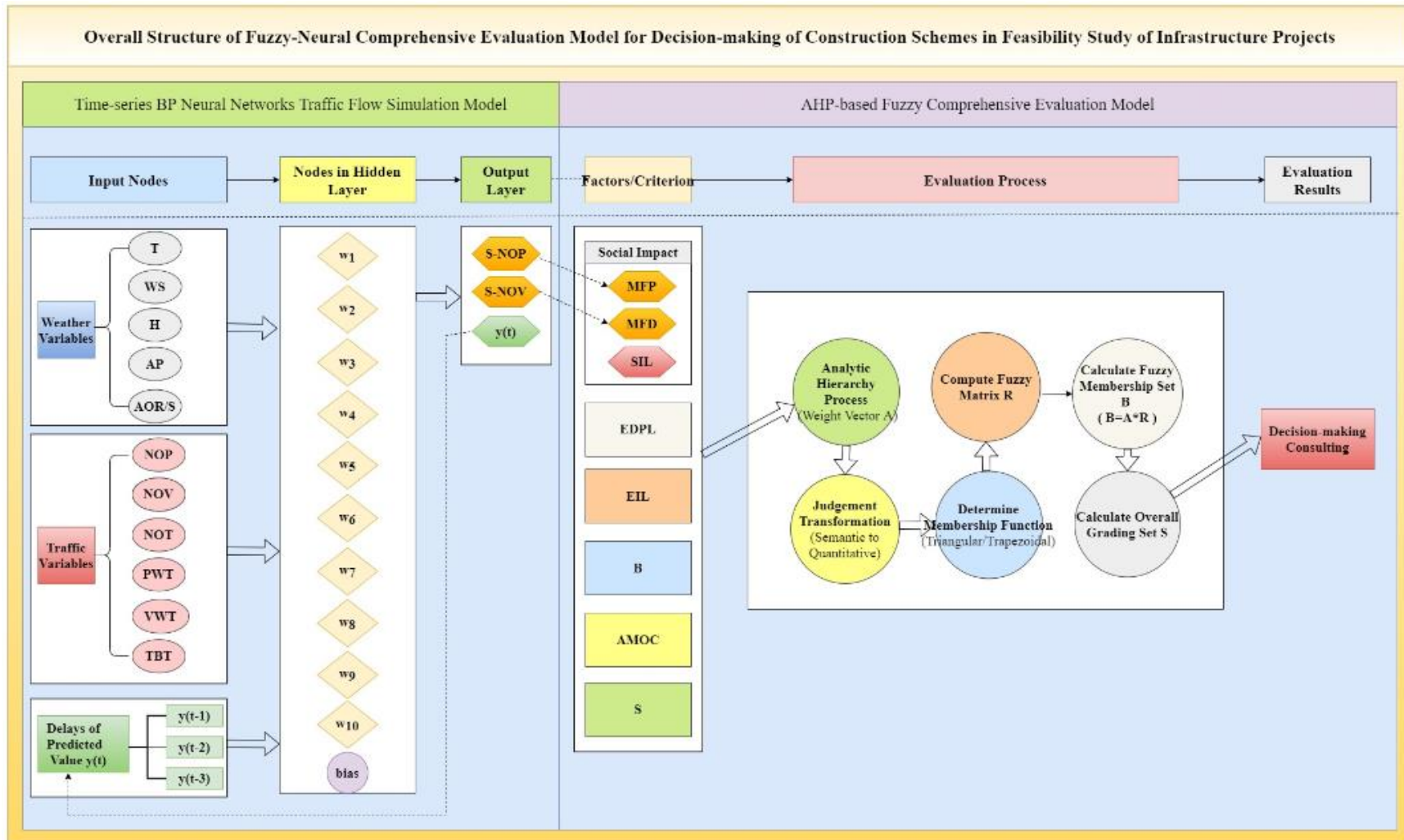
- Vassi, A. & Vlastos, T. (2014). A Review and Critical Assessment of Cycling Infrastructures Across Europe. *WIT Transactions on Ecology and the Environment*, 191.
- Wang, C., Zou, Y. S. and Qu, G. Y. (1999). Grey System Theory in the Selection of Bridge Design Plans. *Journal of Chongqing Jiaotong Institute*, 18(4), 125-130.
- Wen, K.-L. (2008). A MATLAB toolbox for grey clustering and fuzzy comprehensive evaluation. *Advances in Engineering Software*, 39(2), 137–145.
- Wu, C. H., Ho, J. M., and Lee, D. T. (2004). Travel-time Prediction with Support Vector Regression. *IEEE Transactions on Intelligent Transportation Systems*, 5 (4), 276–281.
- Xu, Y. L. (2018). The Shortest Path Analysis Method of Expressway Network Based on Clustering Analysis Algorithm. *Advances in Transportation Studies*, 3, 125–132.
- Yao, Z. S., Shao, C. F., and Gao, Y. L. (2006). Research on Methods of Short-term Traffic Forecasting Based on Support Vector Regression. *Journal of Beijing Jiaotong University*, 30 (3), 19–22.
- Young, G. I. M. (1970). Feasibility studies. *Appraisal Journal*, 38, 376-383.
- Young, P.C., Spear, R.C., Hornberger, G.M. (1978). Modeling Badly Defined Systems: Some Further Thoughts. In: *Proceedings SIMSIG Conference*, Canberra, pp. 24-32.
- Yun, S., & Caldas, C. (2009). Analyzing Decision Variables That Influence Preliminary Feasibility Studies Using Data Mining Techniques. *Construction Management & Economics*, 27(1), 73–87.
- Yuwei Bie, Qiu, T. Z., Can Zhang, & Cunbao Zhang. (2017). Introducing Weather Factor Modelling into Macro Traffic State Prediction. *Journal of Advanced Transportation*, 1–15.

- Zavadskas, E. K., & Turskis, Z. (2008). A New Logarithmic Normalization Method in Games Theory. *Informatica*, 19(2), 303–314.
- Zhang, G., Y. Hu, M., Eddy Patuwo, B., & C. Indro, D. (1999). Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis. *European Journal of Operational Research*, 116(1), 16–32.
- Zhang, H., Gao, S., Zhang, Y., & Yang, F. (2015). Performance Evaluation of the Listed Real Estate Companies in China Based on Fuzzy Neural Networks: The Perspective of Stakeholders. *Journal of Real Estate Practice and Education*, 18(2), 195–215.
- Zhang, W., Lu, J., & Zhang, Y. (2016). Comprehensive Evaluation Index System of Low Carbon Road Transport Based on Fuzzy Evaluation Method. *Procedia Engineering*, 137, 659–668.
- Zhang, Z. S., Guo-Guang, H. E., and Hua-Pu, L. U. (2009). Short-term Traffic Flow Forecasting Based on K-nearest Neighbors Non-parametric Regression. *Journal of Systems Engineering*, 24 (2), 178–183.
- Zhankaziev, S., Gavrilyuk, M., Morozov, D., & Zabudsky, A. (2018). Scientific and Methodological Approaches to the Development Of a Feasibility Study for Intelligent Transportation Systems. *Transportation Research Procedia*, 36, 841–847.
- Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2018). T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *Journal of Latex Class Files*, 14 (8), 1–11.
- Zheng, W. Z., Lee, D. H., and Shi, Q. X. (2006). Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. *Journal of Transportation Engineering*, 132(2), 114–121.

Zhong, K., Song Z., Jain, P., Bartlett, P. L., and Dhillon, I. S. (2017). Recovery Guarantees for One-hidden-layer Neural Networks. The 34th International Conference on Machine Learning, Sydney, Australia.

Ziara, M., Nigim, K., Enshassi, A., & Ayyub, B. M. (2002). Strategic Implementation of Infrastructure Priority Projects: Case Study in Palestine. Journal of Infrastructure Systems, (1), 2.

APPENDIX A: LAYOUT OF THE PROPOSED MODEL SYSTEM IN THIS STUDY



APPENDIX B: CODING SCRIPT FOR 15-MIN NOP/NOV TIME-SERIES BP NEURAL NETWORK SIMULATION

```
% This script assumes these variables are defined:
% InputTrain - input time series.
% OutputTrain - feedback time series.

X = tonndata(InputTrain,true,false);
T = tonndata(OutputTrain,true,false);

% Choose a Training Function
trainFcn = 'trainbr'; % Bayesian Regularization backpropagation.

% Create a Nonlinear Autoregressive Network with External Input
inputDelays = 1:3;
feedbackDelays = 1:3;
hiddenLayerSize = 10;
net = narxnet(inputDelays,feedbackDelays,hiddenLayerSize,'open',trainFcn);

% Choose Input and Feedback Pre/Post-Processing Functions
net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};
net.inputs{2}.processFcns = {'removeconstantrows','mapminmax'};

% Prepare the Data for Training and Simulation
[x,xi,ai,t] = preparets(net,X,{ },T);

% Setup Division of Data for Training, Validation, Testngn
net.divideFcn = 'dividerand'; % Divide data randomly
net.divideMode = 'time'; % Divide up every sample
```

```

net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

% Choose a Performance Function
net.performFcn = 'mse'; % Mean Squared Error

% Choose Plot Functions
% For a list of all plot functions type: help nnplot
net.plotFcns = {'plotperform','plottrainstate', 'ploterrhist', ...
    'plotregression', 'plotresponse', 'ploterrcorr', 'plotinerrcorr'};

% Train the Network
[net,tr] = train(net,x,t,xi,ai);

% Test the Network
y = net(x,xi,ai);
e = gsubtract(t,y);
performance = perform(net,t,y)

% Recalculate Training, Validation and Test Performance
trainTargets = gmultiply(t,tr.trainMask);
valTargets = gmultiply(t,tr.valMask);
testTargets = gmultiply(t,tr.testMask);
trainPerformance = perform(net,trainTargets,y)
valPerformance = perform(net,valTargets,y)
testPerformance = perform(net,testTargets,y)

```

```

% View the Network

view(net)


% Plots

% Uncomment these lines to enable various plots.
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, ploterrhist(e)
%figure, plotregression(t,y)
%figure, plotresponse(t,y)
%figure, ploterrcorr(e)
%figure, plotinerrcorr(x,e)


% Closed Loop Network.

netc = closeloop(net);
netc.name = [net.name ' - Closed Loop'];
view(netc)

[xc,xic,aic,tc] = preparets(netc,X,{ },T);
yc = netc(xc,xic,aic);
closedLoopPerformance = perform(net,tc,yc)


% Multi-step Prediction

numTimesteps = size(x,2);
knownOutputTimesteps = 1:(numTimesteps-5);
predictOutputTimesteps = (numTimesteps-4):numTimesteps;
X1 = X(:,knownOutputTimesteps);
T1 = T(:,knownOutputTimesteps);
[x1,xio,aio] = preparets(net,X1,{ },T1);

```

```

[y1,xfo,afo] = net(x1,xio,aio);
% Next the the network and its final states will be converted to
% closed-loop form to make five predictions with only the five inputs
% provided.
x2 = X(1,predictOutputTimesteps);
[netc,xic,aic] = closeloop(net,xfo,afo);
[y2,xfc,afc] = netc(x2,xic,aic);
multiStepPerformance = perform(net,T(1,predictOutputTimesteps),y2)
% Alternate predictions can be made for different values of x2, or further
% predictions can be made by continuing simulation with additional external
% inputs and the last closed-loop states xfc and afc.

% Step-Ahead Prediction Network
nets = removedelay(net);
nets.name = [net.name ' - Predict One Step Ahead'];
view(nets)
[xs,xis,ais,ts] = preparets(nets,X,{ },T);
ys = nets(xs,xis,ais);
stepAheadPerformance = perform(nets,ts,ys)

% Deployment
% Change the (false) values to (true) to enable the following code blocks.
% See the help for each generation function for more information.
if (false)
    % Generate MATLAB function for neural network for application
    % deployment in MATLAB scripts or with MATLAB Compiler and Builder
    % tools, or simply to examine the calculations your trained neural
    % network performs.

```



```

    genFunction(net,'myNeuralNetworkFunction');
    y = myNeuralNetworkFunction(x,xi,ai);
end
if (false)
    % Generate a matrix-only MATLAB function for neural network code
    % generation with MATLAB Coder tools.
    genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
    x1 = cell2mat(x(1,:));
    x2 = cell2mat(x(2,:));
    xi1 = cell2mat(xi(1,:));
    xi2 = cell2mat(xi(2,:));
    y = myNeuralNetworkFunction(x1,x2,xi1,xi2);
end
if (false)
    % Generate a Simulink diagram for simulation or deployment with.
    % Simulink Coder tools.
    gensim(net);
end

```

APPENDIX C: ONLINE QUESTIONNAIRE ASSOCIATED WITH THIS STUDY



This survey will take you approximately five (5) minutes to complete.

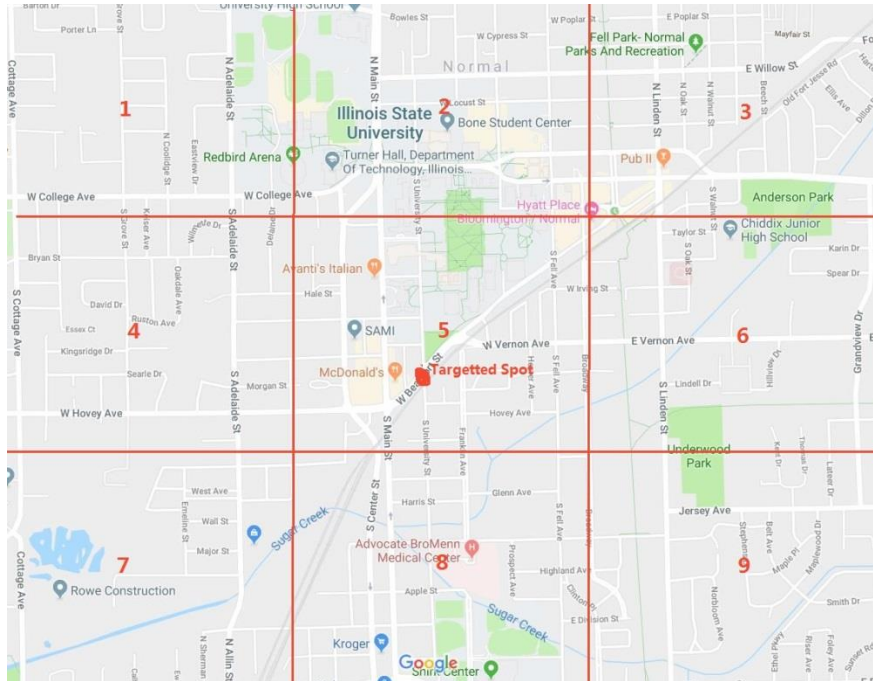
To better understand how individuals, especially ISU students, value a hypothetical pedestrian bridge built at the intersection of University Street and Beaufort Street in Normal, IL, we are asking for your help in answering several questions. A brief description of the hypothetical pedestrian bridge in a feasibility study is below:

The hypothetical pedestrian bridge will be constructed at the intersection of University Street and Beaufort Street in Normal, IL, which is designated to link the separated South University Street by the railroad in order to ease the safety concerns for pedestrians based on an official email from ISU President Larry Dietz about how standing freight trains periodically blocking the rail crossing has created serious delays for pedestrians and motorists alike. Apparently, this issue has a significant impact on the ISU community.

As a thank you for the participation, please sign up for the chance to win following prize package at the end of this survey, which includes:

- Ten randomly selected participants will receive a \$20 gift card to Wal-Mart.

1. What area of the map is your residence located in?



Have you ever been blocked by the moving trains when you hurry up for some events by car or feet? If yes, how did you feel? (Select all that apply)

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
Impatient					
Irritable					
Complaining					
Not surprising					
Heartsease					

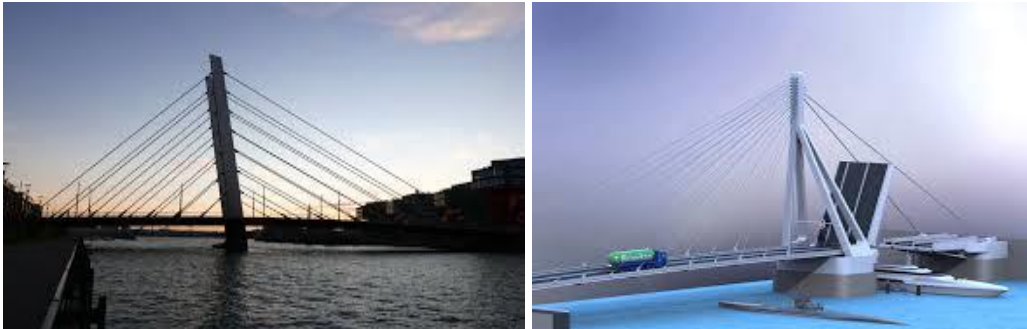
2. How many times have you been blocked by the standing trains weekly?

- a. 0 ~ 5
- b. 5 ~ 10
- c. 10 ~ 15
- d. 15 ~ 20
- e. More than 20

3. Generally, how long do the cargo or passenger trains delay you when you wait for access to the other side of the railway?
- a. 0 ~ 5 min
 - b. 5 ~ 10 min
 - c. 10 ~ 15 min
 - d. 15 ~ 20 min
 - e. More than 20 min
4. Based on the inconvenience, how urgent do you think a pedestrian bridge should be built at the intersection mentioned above?

Strongly	Comparatively	Mildly	Little	Not at all

5. If it is proper to get the pedestrian bridge built, which kind of bridge is your favorite?
- a. **Cable-stayed bridge** (Bridge deck is supported by the cables from one or more towers)



- b. **Truss bridge** (A bridge whose load-bearing superstructure is composed of a truss)



- c. **Suspension bridge** (A bridge in which the deck is hung below suspension cables on vertical suspenders)



- d. **Arch bridge** (A bridge with abutments at each end shaped as a curved arch)



6. In terms of the benefits, this pedestrian bridge may have, please select the applicable items.
- a. Improve safety for pedestrians and bicyclists, especially the ISU community, by reducing the risky “beat the train” behavior.
 - b. Reduce the delay time caused by standing long cargo trains.

- c. Promote the walkability for a specific proportion of ISU students to and from campus.
- d. Stimulate economic development in the surrounding area by landmark effect.
- e. Others

7. If taking environmental sustainability into account, what key sustainable points this bridge should have according to your perspective?

- a. Utilizing eco-friendly, resilient materials, and sustainable design to reduce the maintenance cost in the future.
- b. Installing a Solar-panel roof on this bridge to generate and provide electricity for night lighting and network surveillance equipment.
- c. Reducing the emission of carbon dioxide by promoting more ISU students in the area to walk to and from campus.
- d. Others.

8. Do you have any other thoughts or comments regarding the needs of building this pedestrian bridge?

Survey Complete

We highly appreciate your time and consideration in completing this survey. Through your help, we can collect the needed information to make this study more brainstorming. Besides, your mind and idea will contribute a lot to this feasibility study of a pedestrian bridge for rail and road

crossing. If you have any questions or concerns, please feel free to contact Xi Hu from the Department of Technology at Illinois State University at xhu1238@ilstu.edu.

If you would like to enter winning the prize package, please provide your contact information.

This information will not be used for any other purposes.

Name	
Email Address	