

Modelling and Simulation of Traffic Dynamics in Almaty City

Ayala Amir

Satbayev University, Almaty, Kazakhstan
E-mail: amirova.ayala@gmail.com

Abstract—This work examines concerns associated with traffic analysis applicable within a given region of Almaty City employing real data and modelling approaches. In light of analyzing traffic congestion and accidents, the study unveils the following caveats and research limitations: In order to overcome the above challenges, the study proposes the following ways: The study also has the following achievements: The solution suggested to address the problem is adopting an Adaptive Transport Control System to improve traffic management. By employing simulation models and real data the study aims to enhance traffic flow and the transport system in the Almaty City. The research utilizes SUMO an open-source traffic simulation software to perform simulation settings with consideration of parameters like signal lights, separation of lanes, and probability for appearance of vehicles. Also, in this study, parameters of BPR function are calibrated to improve the modeling of traffic flow in order to support the application of efficient traffic control measures. The analysis shows that even minor changes to the BPR parameters can lead to a dramatic increase in the model's accuracy, which will be useful for understanding the planning of cities. Subsequent work is to conduct studies on these models across diverse datasets and road types, as well as incorporating real-time data to the models, and covering more extensive facets relating to the traffic flow model so as to provide paramount support in enhancing the efficiency of urban transportation systems.
Keywords: *Simulation Model, Transport Network, Traffic Signal Controller, Transportation System, SUMO, Traffic Flow Optimization, Adaptive Transport Control System, Real-Time Data Collection, BPR Function, Urban Traffic Management*

I. INTRODUCTION

In today's modern world, a pressing issue revolves around the rapid proliferation of vehicles in urban environments. This surge has led to various negative outcomes, including health problems, air pollution, and significant travel delays due to traffic congestion. Efficiently managing traffic flow has become a critical challenge. This article aims to address this challenge by focusing on optimizing traffic fluidity in Almaty, a city facing these issues. To achieve this goal, we present a detailed model focusing on a specific territory within Almaty.

To manage traffic flows, the first order is to manage traffic flow in a single intersection, it is important because if the system works at a single intersection therefore, it is

the assumption that system will work in multi-intersections. Second, would be to conduct experiments for multi-intersections, then to manage traffic flows in the territory of city Almaty, as shown in Fig. 1, with more than 100 intersections using real data. Kurmankhojayev et al. [1] considered the problem of adaptive light control at a single isolated intersection. They report the queue length reduction under the stationary demand condition as well as, our aim is to use their model in our network. In every research, data plays a major role, and this paper shows real data that was collected by 2GIS, which contains information about addresses, traffic congestion and accidents.

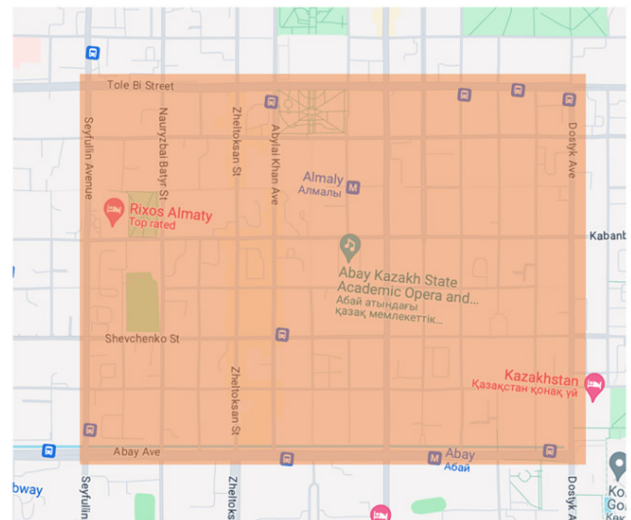


Fig. 1: Target Area: Square from Tole bi Avenue to Abay Avenue and from Seifullin Avenue to Dostyk Avenue

We have completed a simulation model in the territory of city Almaty and completed data collection of this area.

The rest of this paper is organized as follows:

- Simulation model is discussed in section II.
- In section III, Data collection is interpreted.

Finally, conclusions are stated in section IV.

II. LITERATURE REVIEW

Traffic flow modelling has evolved significantly with the advancement of simulation tools such as SUMO (Simulation of Urban Mobility) developed by the German Aerospace Centre (DLR). SUMO is an open-source, highly

portable, and microscopic traffic simulation package designed to handle large road networks. Foundational theories of traffic flow, as discussed by [2] and [3], provide the basis for vehicle dynamics and behaviour modelling. Modern methodologies leverage tools like SUMO for more detailed and scalable simulations.. [4, 17, 18] introduced SUMO's capabilities, highlighting its flexibility in simulating urban traffic scenarios. Further, [5,16] provided an overview of SUMO's applications, demonstrating its effectiveness in various research contexts. The integration of SUMO with other traffic assignment models has been explored by [6,13,14], showcasing its potential for comprehensive traffic management solutions. Additionally, Erdmann [7,15] discussed future advancements and ongoing developments in SUMO, emphasizing its role in future mobility simulation. Case studies specific to Almaty, such as those by [9-12], offer practical insights into local traffic management, suggesting that the integration of SUMO can enhance the accuracy and efficiency of traffic flow models tailored to Almaty's unique urban environment. Despite these advancements, there remain gaps in localized modelling and the incorporation of emerging technologies, which are essential for developing more effective traffic management strategies.

III. SIMULATION MODEL

In our work to make simulations we use SUMO DLR. SUMO transport network is a directed graph. It consists of roads and intersections. The intersections are represented by nodes (junctions) and the roads by unidirectional edges.

There are two methods of building transport network:

- *.osm(file taken from OpenStreetMap)
- by hand

OOSM gives opportunity to get *.osm file which we can convert into SUMO *.net file which helps us to model the network. However, there are some disadvantages of using this method.

- (a) Hard to control and to make changes into the *.net file
- (b) Lots of unnecessary data such as: buildings, unknown ids of the streets.

In the second method, human observers were utilized to measure and verify road lane configurations. This process involved several key steps:

- *Distance Measurement Using 2GIS*: Human observers used the 2GIS application to measure the distance on each segment of the road. This application provides accurate mapping and measurement tools, allowing observers to determine the length of each road section precisely.
- *Lane Identification Using Maps*: After measuring the distances, the observers used detailed maps to examine the number of lanes on each road segment. The maps provided by Yandex Maps offer high-resolution imagery and detailed information

about road layouts, which facilitated the accurate identification of lane configurations.

- *Field Verification*: To ensure the accuracy of the data collected through 2GIS and maps, the observers conducted field verification. This involved physically inspecting the road segments in question to confirm the number of lanes and other relevant features. This step was crucial for validating the information obtained from digital tools and ensuring its reliability.

By combining digital measurements with field verification, this method ensured a high degree of accuracy in identifying and recording the number of lanes on various road segments. The integration of technology and direct observation provided a robust framework for the transport modelling.

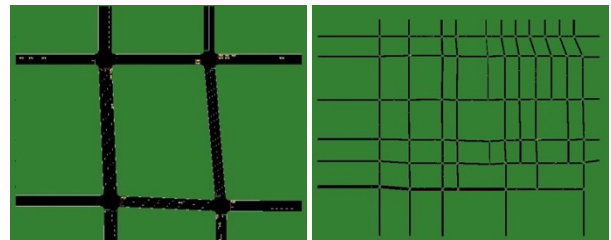


Fig. 2: Test Transport Network

As shown in Fig. 2, there is shown test transport network modeled in SUMO DLR. Target area is “square” from Tole bi Avenue to Abay Avenue and from Seifullin Avenue to Dostyk Avenue. It contains 100 nodes in which 49 of them signalized that means there located the traffic light and 51 of the nodes are unregulated, without traffic lights. As well as test network consists of 202 edges.

During our project, several inquiries arose:

- *Challenges in Obtaining Signal Plans for the Entire Transport Network*: We investigated the potential difficulties associated with acquiring signal plans for the entire transportation network. To assess the operational dynamics of the transportation system within our targeted area, we procured a Signal Plan. Our assistants meticulously observed traffic lights at each intersection to gather this data. The process of data collection spanned a period of two weeks. However, a significant obstacle emerged – determining the offset time between intersections posed the greatest challenge during the experiment.
- *Enhancing Efficiency in Official Service Provision*: We contemplated strategies to expedite more efficient service provision by authorities, particularly concerning the delivery of service plans. There exists a possibility of accessing real-time data in a readily usable format, which could streamline the process of obtaining signal plans and calculating offset times. Additionally, it's noteworthy that signal plans in the city undergo bi-

daily changes. For this study, we solely relied on the morning signal plan, which spans from 5 a.m. to 8 p.m.

- *Current Methods Employed:* The prevailing solution entails data collection solely through human observation and the physical calculation of offset times.
- *Status of Data Collection:* The data collection process remains incomplete. The primary hindrances include the absence of an evening signal plan, imprecise offset time calculations, and challenges in observing certain intersections.

IV. DATA COLLECTION

A. From 2GIS

This section discusses the methodology employed to gather data from 2GIS, an international Geo Information System company offering mapping services and traffic-related information. 2GIS provides real-time data on traffic congestion and average speeds on specific road segments, with a claimed accuracy of 95 percent. This data is periodically verified by GIS specialists and updated through a contact centre. Access to 2GIS data is primarily restricted.

To collect data on average speeds, we utilized a script written to extract information from 2GIS. We employed JQuery Ajax programming for this task, which allowed us to interface with the 2GIS application and retrieve real-time average speed data. Our objective was to automate the data collection process to enable continuous storage of information.

2GIS categorizes traffic congestion into six levels, ranging from level 1 indicating no traffic and high average speeds, to level 6 indicating significant delays and lower average speeds. The average speed data is updated every 5 minutes. Initially, we collected average speed data over a two-week period at 20-minute intervals. Each road segment is identified by a unique ID represented by a polygon containing coordinates on the map, with the corresponding average speed displayed, as shown in Fig. 3.



Fig. 3: Data on Average Speeds that Appear Using the Extension JQuery Ajax

For research purposes, we took the “square” of the streets of Almaty, starting from Tole bi Avenue to Abai

Avenue and starting from Seifullin Avenue to Dostyk Avenue, as shown in Fig. 1.

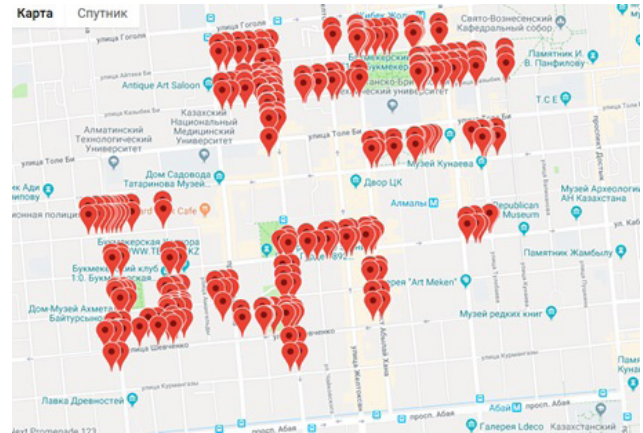


Fig. 4: The Created Map on Google Maps

Next, we wrote a script and automated the program so that every 20 minutes *.txt files with real-time data were saved. Data collection started on March 17 at 3 a.m. and finished on April 1 at 3 a.m. As a result, 1098 data files on average speeds were collected in 2 weeks. Subsequently, these data as shown in Fig. 4 can serve as input data for counting the number of vehicles at the intersections.

B. Average Speed to Number of Vehicles

First, we collect the data. To do that we run the simulation with different parameters. Each simulation lasts for 50 000 seconds (simulation time units). We record the simulation states each 6 seconds. The tracked features are the following:

1. Number of vehicles in North-South direction
2. Number of vehicles in South-North direction
3. Number of vehicles in West-East direction
4. Number of vehicles in East-West direction
5. Average delay in North-South direction
6. Average delay in South-North direction
7. Average delay in West-East direction
8. Average delay in East-West direction
9. Green phase duration in North-South direction (it is the same for South-North)
10. Green phase duration in West-East direction (it is the same for East-West)
11. Current phase index
12. Remaining time (in seconds) till the end of the phase
13. Average speed of vehicles in North-South direction
14. Average speed of vehicles in South-North direction
15. Average speed of vehicles in West-East direction
16. Average speed of vehicles in East-West direction

We choose between 256 demand models and 9 signal plans. It results in 2304 simulations.

We do not need to annotate the data for the average speed estimation task because we ourselves generated the demand models, hence, their parameters are known to us.

The average speed estimation model (NN) is trained on this data.

1) Demand Models

Our demands are based on the second based probabilities of vehicles being generated. For instance, if it is 0.1 in North-South direction then it means that the probability of a vehicle being generated in this direction each unit time is 0.1, as shown in Table 1.

Table 1: Demand Models in Terms of Simulation Step-Based Probability of Vehicles Being Generated in Specified Directions

North-South	South-North	West-East	East-West
0.1	0.1	0.1	0.1
0.1	0.1	0.1	0.2
0.1	0.1	0.2	0.1
0.1	0.2	0.1	0.1
...	...	...	...
0.4	0.4	0.4	0.4

2) Signal Plans

We assume that the intersection is a two-phase intersection. We have limited ourselves only to controlling the green phase duration in North-South and West-East directions are shown in Table 2, i.e., we fixed the duration of the amber phase.

Table 2: Defined List of Signal Plans. The Numbers Represent the Durations of Green Phases in the Specified Directions in Seconds

North-South and South-West	West-East and East-West
12	12
12	18
...	...
24	24

C. Average Speed Estimation

1) Neural Network Method

To predict demand using the collected data, a fully connected feedforward neural network (NN) is employed. The model takes as input a 60-dimensional vector, obtained by concatenating five consecutive records (each comprising 12 entries) from the collected data and flattening it into a single vector. The output of the model is a one-dimensional vector containing estimates of average vehicle speeds in each direction.

The architecture of the model is determined through a series of experiments, starting from a basic structure with one hidden layer and progressively increasing in complexity as necessary. Initially, the number of hidden neurons is calculated using a heuristic known as the “rule of thumb,” which serves as a starting point for designing the neural network architecture.

$$N_{hn} = \frac{N_i + N_o}{2}$$

where the number of hidden neurons, denoted as N_{hn} ,

relates to the number of neurons situated in the hidden layer of a neural network. Here, N_i represents the number of neurons in the input layer, and N_o denotes the number of neurons in the output layer.

D. BPR Function

Common formulations of volume delay functions rely on the standard Bureau of Public Roads (BPR) function, which has been widely adopted in US guidelines and transportation planning practices since its inception. However, studies have observed limitations with default BPR functions and their modifications, noting potential overestimation of travel times during congested conditions and overestimated traffic volumes when capacity is exceeded. Conversely, these functions may underestimate traffic flows during relatively free-flow conditions. Adjusting BPR parameters can lead to more consistent travel times in uncongested traffic but increases sensitivity to volumes nearing or exceeding capacity.

Empirical evidence suggests that the steepness of the speed-volume curve in congested conditions varies between road types, such as freeways and urban arterials, and is strongly influenced by road design.

For formulating the problem, the following notation will be used:

- q : flow (vehicles per hour)
- v : speed (kilometers per hour)
- v_0 : free-flow speed (kilometers per hour)
- q_{max} : capacity (vehicles per hour)
- a, b : estimated parameters of the speed-flow function

To illustrate the method proposed in the paper, we will utilize the simplest BPR function, formulated with

$$\hat{t} = t_0 \left(1 + a \left(\frac{q}{q_{max}} \right)^b \right) \quad (1)$$

$$\frac{\hat{t}}{t_0} = \left(1 + a \left(\frac{q}{q_{max}} \right)^b \right) \quad (2)$$

By assuming constant length, the relative speed drop can be derived from BPR function (1) as inverse of travel times:

$$\hat{v} = v_0 \left(1 + a \left(\frac{q}{q_{max}} \right)^b \right) \quad (3)$$

In the paper, we discuss the estimation problem of the BPR function where we look for optimal values of parameters a and b . The objective is to reproduce the observed data with BPR function; that is, the modelled speeds should match the measured ones, which can be formalized with the following least-squares formula computed as the sum of squared difference between modelled and observed speeds for all the observations τ in the dataset T :

$$\frac{1}{n} \sum (|\hat{v} - v|)^2 \quad (4)$$

We analyze the solution of problem (4) itself, since it is straightforward and can be solved with any linear solver (we formulated the problem for the least-squares method).

E. Different Usage of Average Speed

- *Predicting Road Conditions:* Monitoring the average speed on a road segment can serve as an indicator of traffic congestion. A decrease in average speed typically signals the onset of a traffic jam, enabling pre-emptive measures to mitigate severe congestion and its associated consequences.
- *Vehicle Counting:* Average speed data in each direction can be utilized to estimate the number of vehicles at intersections. This data can subsequently inform the control of traffic lights, as demonstrated in the work of Kurmankhojayev et al. [1], which addressed adaptive traffic light control at a single isolated intersection.
- *Estimating Average Travel Time:* To assess road traffic congestion, one can measure the Annual Average Travel Time. By utilizing the known length of the road segment and the average speed, it is possible to calculate the average travel time, providing a useful metric for traffic analysis.

V. EXPERIMENT AND RESULTS

An isolated intersection built in SUMO DLR [10]. We set the intersection geometry as following, as shown in Fig. 5:

- Length of an edge is 100 meters
- An edge has 2 lanes
- Average vehicle's length is 5 meters
- Total number of halting vehicles is in the range $0 \leq N \leq 80$

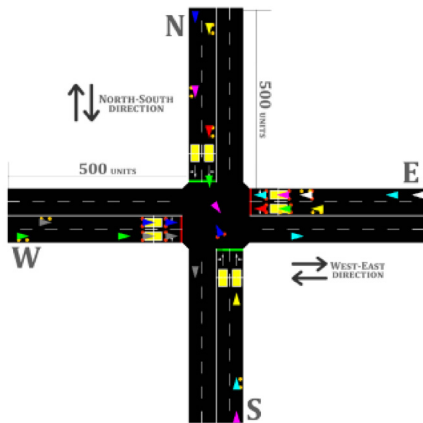


Fig. 5: Test Intersection

The problem was simplified by excluding the following:

- U-turns, public transport flows, pedestrian flows, traffic rules violation, traffic road accidents, and parking.

- We assume that the acquired data are free of noise and available on request.

F. Fully Connected Neural Network (FCNN)

The architecture of our Fully Connected Neural Network (FCNN) model is summarized in Table 3. The model comprises four layers: an input layer, two hidden layers, and an output layer. The input layer consists of 60 neurons, which correspond to the 60-dimensional input vector derived from the collected data. The first hidden layer contains 30 neurons, while the second hidden layer has 44 neurons. Finally, the output layer consists of 4 neurons, which represent the predicted outcomes. This architecture was determined through a series of experiments aimed at optimizing the model's performance for traffic flow prediction. The choice of the number of neurons in each layer was guided by the need to balance model complexity and computational efficiency.

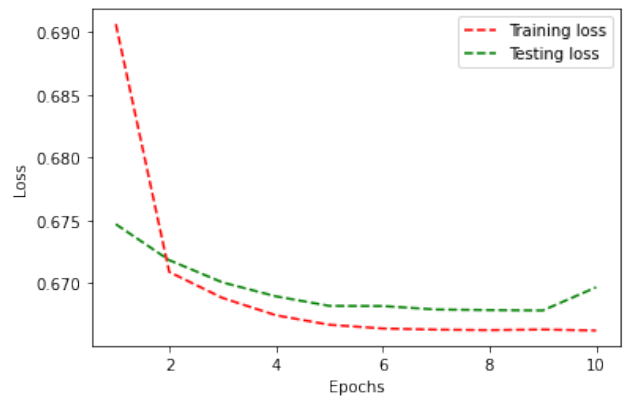


Fig. 6: The Loss Values of Training and Testing the FCNN Based Model (the Loss is Calculated using the MSE Function)

The graph, as shown in Fig. 6, the loss curves for both the training and testing datasets.

Training Loss (red dashed line): The loss values of the function on the training data decreases with epochs showing that the model is learning from the data. However, the loss values are still comparably high, which makes a question about the ability of the model to identify the patterns of the given data.

Testing Loss (green dashed line): The values of the loss function on the testing dataset also decline at first, but then begin to slightly grow in some number of Epochs. The testing loss is high almost every epoch, which means that the model is not able to generalize on different data set.

In total, the findings show that Fully Connected Neural Network cannot capture the mapping between averaged speed or the number of vehicles. All the values of loss are high and remain high for both the training and testing datasets, which might indicate that the model is too simple, or not well optimized for the given problem. This could be due to the problems of both *prima facie* insufficient model complexity as well as insufficient features engineering as well as suboptimal hyperparameters selection.

G. BPR Function

Our experimental results yielded promising outcomes:

Training Loss: When utilising the fine-tuned BPR function, the model had a training loss of 0.013, thereby suggesting a good fit to the training data.

Testing Loss: The testing loss was 0.014, Thus our model generalised well with unseen data, thereby overcoming the drawback of BPR function.

Optimized Parameters: The evaluation of the parameters “a” and “b: ” yielded the best-fit figures of 6.89×10^{-6} and 3, respectively. These values are different from those estimated in the traditional manner and contributed to the refinement of our model to a considerable extent.

All the adjusted function parameters of BPR function improved the accuracy of modelling traffic flow issue. With these parameters adjusted it proved possible to reduce the over-estimated travel time during traffic jams and obtain more accurate traffic volume estimates when the road space is fully utilised. This improvement is a major step towards better control of the traffic flow and better design of roads and related structures.

VI. CONCLUSION

In conclusion, our study demonstrates the potential for enhancing traffic flow modelling by fine-tuning the parameters of the BPR function. The standard BPR function, while valuable, can be significantly improved to provide more accurate predictions across various traffic scenarios. These improvements have practical implications for transportation planners and researchers, offering a path toward more precise and effective traffic management strategies in congested urban environments. Future research should focus on validating these findings across different datasets and road types, integrating real-time data, and expanding the scope of traffic flow models to include a broader range of influencing factors. By continuing to refine and validate these models, we can achieve more efficient and responsive urban transportation systems, ultimately reducing congestion and improving urban mobility.

A. Why do the Results Matter?

The implication of the finding in our study is crucial since it shows signs that ‘tuning’ of the BPR function parameters enhances the accuracy of the traffic flow models. This assertion can result in making much more accurate prediction of travel time and traffic flow, which are very vital for coordinating of traffic and for planning of transport structures. Valid, reliable and achievable traffic flow models are important for alleviation of traffic congestions, tuning of traffic signals and enhancement of city accessibility. In allowing for more accurate transportation planning, our research outcomes will assist urban planners and policy-makers to effectively design measures that will decrease traffic congestion and optimise the efficiency of the urban transport systems.

Finally, one drawback of the results is what the results can’t tell us.

This brings the final conclusion that although the results mirror certain elements of success our approach does not offer a solution of general application to virtually any traffic scenario and road types. This work was based on a particular type of data collected in a particular environment of an urban city and therefore the findings and their application may not be used for other cities or even on different road networks. Further, even though, we have enhanced the BPR function to increase the increased accuracy of results for our dataset, there could be other factors that affect the traffic circulation yet disregarded by our model including drivers’ behavior and weather conditions among others.

What specific industry activities or empirical research should ensue?

To build on our findings, the following practical actions and scientific studies are recommended:

- *Further Validation:* Further testing of the adjusted BPR function parameters using other datasets from other cities and different roads is required. This will not only make it easier for us to generalize out finding but also help see the extent of its general applicability.
- *Real-Time Data Integration:* Examine aspects of real-time traffic information for improvement of models adjusting the traffic flow, with the focus on the dynamic traffic management systems.
- *Traffic Light Control Optimization:* Use the results from our model to plan, install, and then analyze the performance of adaptive traffic light control systems on the road.
- *Comprehensive Modeling:* It would be useful to introduce other parameters including the behavior of drivers, weather factors, and certain incidents to determine a greater amount of influences on traffic flow.

While working on this piece the Amir A. resorted to the use of ChatGPT to enhance language and readability. Amir A. cleans up the content after using this tool and bears total responsibility for the content of the publication.

REFERENCES

- [1] D. Kurmankhojayev, N. Suleymenov, and G. Tolebi. 2017. *Online model-free adaptive traffic signal controller on an isolated intersection*. In 2017 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON).
- [2] Treiber, M., & Kesting, A. (2013). *Traffic Flow Dynamics: Data, Models and Simulation*. Springer.
- [3] Helbing, D. (2001). “Traffic and related self-driven many-particle systems”. *Reviews of Modern Physics*, 73(4), 1067-1141.
- [4] Krajzewicz, D., Hertkorn, G., Rössel, C., & Wagner, P. (2002). “SUMO (Simulation of Urban MObility); An open-source traffic simulation”. *Proceedings of the 4th Middle East Symposium on Simulation and Modelling (MESM2002)*, Sharjah, United Arab Emirates.

- [5] Behrisch, M., Bieker, L., Erdmann, J., & Krajzewicz, D. (2011). "SUMO-simulation of urban mobility: An overview". Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation.
- [6] Gawron, C. (1998). "Simulation-based traffic assignment". Proceedings of the 2nd SUMO User Conference.
- [7] Erdmann, J. (2015). "SUMO's Roadmap for Future Mobility Simulation". Proceedings of the SUMO User Conference 2015
- [8] Kalimova, A., & Myrzakhmetov, B. (2019). "Urban Transport System and Traffic Management: The Case of Almaty". Procedia Engineering, 178, 69–75.
- [9] Bakhtiyarov, M., & Kurmanbekov, A. (2020). "Evaluation of the effectiveness of traffic management measures in Almaty". Transportation Research Procedia, 47, 113–120.
- [10] G. Tolebi, N. Dairbekov, and D. Kurmankhojayev. 2020. *Link Flow Estimation on an Isolated Intersection Based on Deep Learning Models*. In 2020 International Review of Automatic Control (I.R.E.A.CO.).
- [11] Singh, G., Chaudhary, N., Ather, D., Kler, R., & Arora, M. (2023, December). Revolutionizing Remote Healthcare: Proposing An IoT & Arduino-Based Integrated Approach for Real-Time Health Monitoring. In 2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 1-5). IEEE.
- [12] Arora, M., Singh, G., Ather, D., Chaudhary, N., & Kler, R. (2023, December). Forecasting Inbound Tourism in Uzbekistan: Leveraging AI and ARIMA Models for Economic Growth Insights. In 2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 1-4). IEEE.
- [13] Ather, D., Chaudhary, N., Singh, G., Beig, T., & Kler, R. (2023, December). Enhancing Used Automobile Valuations: A Data-Cleaning and Linear Regression Approach for Predicting Prices in Competitive Market. In 2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 1-5). IEEE.
- [14] E. Cascetta. *Transportation Systems Analysis – Models and Applications*, volume 29 of *Springer Optimization and Its Applications*. Springer, 2009.
- [15] R. P. Roess, E. S. Prassas, and W. R. McShane. *Traffic Engineering*. Prentice Hall, 3rd edition, 2004.
- [16] J. Barceló, editor. *Fundamentals of Traffic Simulation*. Springer, 2010. DOI: 10.1007/978-1-4419-6142-6.
- [17] L. C. Bazzan and F. Kluijgl. 2014. *A review on agent-based technology for traffic and transportation*. Knowledge Engineering Review 29, 3 (2014), 375–403.
- [18] Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. 2012. *Recent Development and Applications of SUMO - Simulation of Urban MObility*. International Journal On Advances in Systems and Measurements 5, 3&4 (December 2012), 128–138