Proceedings of the 2nd International Conference on Civil Engineering Fundamentals and Applications (ICCEFA'21)
~~Seoul, South Korea~~ Virtual Conference - November 21 - 23, 2021
Paper No. 134
DOI: 10.11159/iccefa21.134

Modelling of Delay for Protected/Permitted Left Turning Vehicles using Multigene Genetic Programming

Nemanja Dobrota[1], Aleksandar Stevanovic[1]
[1]Department of Civil and Environmental Engineering/University of Pittsburgh
3700 O'Hara St, Pittsburgh, PA, USA
ned47@pitt.edu; stevanovic@pitt.edu

Abstract - Vehicular delay represents one of the fundamental traffic signal performance measures. In the past, number of delay models were developed mainly to estimate delays for exclusive phases (movements). In cases of left-turn movements that are served in protected/ permissive mode, there are very few models that can be used to estimate delays. Previously developed models for left/turn protected/permissive mode are based on a number of assumptions related to vehicular arrival/departure patterns. Thus, when used to estimate delays in a real-time manner, such models are prone to erroneous estimates. In this study, to overcome the limitations of current models, authors proposed a novel delay model for protected/permitted left turn operations based on Multigene Genetic Programming (MGGP) technique. Relevant data were collected on a cycle-by-cycle basis using the microsimulation model of real-world arterial. Using the MGGP, a novel delay model and its analytical formulation were proposed and compared with the benchmark model from the literature. The results indicate that the proposed model is more accurate and reliable and should be used as an alternative to traditional models. To strengthen the conclusions of our study, future work is mainly related to the expansion of the utilized dataset used for model development based on the field data. It is imagined that such an expanded dataset and additional options within MGGP will be explored to develop a more robust delay model.

***Keywords*:** *Traffic Signal Performance Measures, Delay, Multigene Genetic Programming, Protected/Permitted Left Turns*

1. Introduction

Modelling of traffic operations at signalized intersection consists of two main components, characteristics of vehicular arrival/departure patterns and duration of signal timing parameters. Interaction between these components is depicted through the various traffic signal performance measures. Initially, traffic operations at signalized intersections were modelled as queuing systems [1]. Formulation of traffic operations as a queuing system enabled the derivation of a number of analytical models for signal performance measures [2,3]. Until nowadays, such models are widely used by current practice for traffic signal performance assessment in the various decision-making processes [4,5].

Perhaps the most important traffic signal performance measure is vehicle delay [6]. In the past, measurements of delay in the field were time-consuming and labour-intensive practices. Therefore, researchers developed delay models that provide delay estimates for limited input data (volumes, signal timing parameters). Starting from the first deterministic delay models [1,2], it was noticed that estimates are not so realistic due to the underlying model assumptions (i.e. uniform vehicular arrivals). Series of following studies aimed to account for stochasticity in vehicular arrivals [7]. The most cited stochastic delay model in literature is the famous Webster model [3]. Soon after, delay models were questioned for accuracy in oversaturated conditions. Many delay models were developed to account for both under and oversaturated conditions [8,9]. Such models showed good promises for delay estimations in so call time-dependent traffic regimes. To overcome the need for various progression factors, oversaturated conditions and to account for specific traffic operations (e.g., during protected/permitted phasing, permitted right-turns-on red), Strong and Rouphail (2006) proposed Incremental Queue Accumulation (IQA) model [10].

The IQA model can be used to estimate delays on the basis of distinctive queuing polygons identified during the cycle as a consequence of arrival/departure flow rate changes [10]. Such polygons are essentially defined for each specific case of traffic signal operations (e.g., protected/permitted) based on which delays are estimated. Previous studies show superior results of IQA when compared to the traditional Highway Capacity Manual 6th edition [11] model [12,13]. However, use of

IQA model becomes very impractical when delay estimates are conducted in real-time manner on cycle-by-cycle basis, since each cycle can have distinctive vehicular arrival/departure patterns; therefore, the shape of polygons needs to be identified for each cycle appropriately. Such a problem becomes especially apparent in the case of left-turning vehicles that are served in protected/permissive mode (more details are given in the following chapter). In particular, IQA and similar analytical models [14] are not recognizing cases when vehicles during the permissive phase arrive and stop (when the opposing queue was completely cleared) due to random or late platoon arrivals from opposing flows. Delays for such vehicles are not captured by existing models and therefore, estimates that are resulting from these models are inaccurate.

Very few scholars tackled the issue related to modelling of delays for protected/permitted left turn movements [15,16]. However, previous work was based on several assumptions regarding characteristics of opposing flows (served as platoons [15]) or developed models were evaluated using the numerical approach (i.e., without evaluating models using field or field-like data from microsimulation [16]). Therefore, the accuracy of the proposed models is still questionable. One approach to model delays is to rely on various machine learning methods. By gathering the knowledge from various observations from the realistic traffic data, ML models are proven to be viable tools to capture non-linearity between input parameters that cause certain delays [17]. From various algorithms so far authors used, ANN [18], kNN, SVR, RF, and XGBoost [19], Multi Gene Expression programming [17]. However, most of the previous studies explicitly considered exclusive through movements, and thus proposed models are not applicable for permitted/protected operations.

Therefore, to overcome these limitations, we use a relatively recent ML technique called MultiGeene Genetic Programing [20]. Authors of this study use the MGGP for two main reasons, (i) it was previously shown that MGGP can outperform widely used machine learning techniques (e.g. Artificial Neural Networks) when used to solve similar problems [17] and (ii) compared to other ML techniques that operate as black-boxes, MGGP outputs analytical model that can be used by both, practitioners and researchers. This study has a twofold contribution, the genetic programming approach is demonstrated for delay modelling, and an analytical formulation for delay based on real-time (i.e., cycle-by-cycle data) is proposed.

2. Delay models for protected/permitted LT

2.1. Incremental Queue Accumulation model for PPLT

The IQA delay model [10,11] estimates queue lengths (Q_q, Q_r) and consequently delays (as the area underneath formed polygon), by using distinctive arrival flow rates during red (q_r) and green (q_g) and two distinctive saturation flow rates during the protected part of the green (s_l) and during permitted part (s_{lt}) of the green as illustrated in Figure 1. Depending on the intensity of opposing traffic flow, the permitted green period consists of a "red" period labeled as shaded red/green portion of g_{per} (when the opposing flow is discharging) and "green" period labeled as shaded green (when opposing flow is fully discharged and vehicles from left-turning lane are utilizing accepted headway to conduct their turn maneuver). Several assumptions are made within the IQA model parameters. The assumption regarding saturation flow rate (s_{lt}) is that vehicles from opposing flow arrive in a random manner [11]. Another assumption is related to the number of sneakers (vehicles that essentially depart from the left-turning lane after the green interval ends) that need to be made in advance and its constant for each cycle [11]. Finally, when calculating the arrival flow rates during green and red (q_g, q_r) it is assumed that vehicles arrive randomly during this period [11]. It needs to be mentioned that all three assumptions are violated in cycle-by-cycle delay estimation procedures when the impacts of nearby coordinated signals are present.

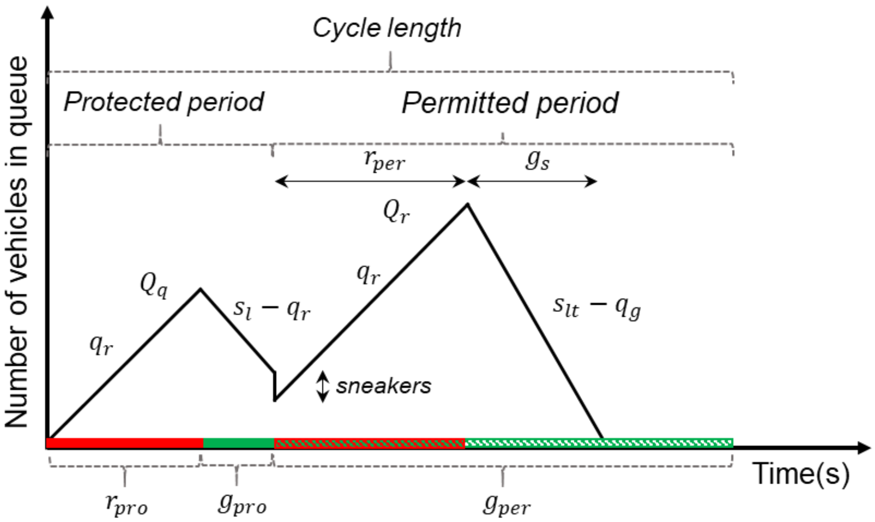

Fig. 1: Incremental Queue Accumulation Polygon for Protected/Permitted Left Turns (adopted from HCM2016 [11]).

Estimation of delay within IQA starts with estimation of queue length Q_* (Eq. 1) during particular duration t_* (s) of distinctive polygon, where $t_* \in \{r_{pro}, g_{pro}, r_{per}, g_s\}$ for particular saturation flow rate $s*$ (veh/s), where $s_* \in \{s_l, s_{lt}\}$ and flow rates q_*(veh/s), where $q_* \in \{q_r, q_g\}$ for N number of lanes. Further, the average delay for the cycle is calculated as the summation of delay with each polygon divided by total volume q (veh/cycle) as shown in Equation 2. It needs to be stated that flow rates represent the ratio of the number of vehicles arriving during green or red (*AoG*, *AoR*) divided by the duration of green or red. For calculations of saturated flow rates during distinctive portions of the cycle, for the sake of brevity, the reader is referred to the source manual [11].

$$Q_* = Q_{*-1} - \left(\frac{s*}{3{,}600} - \frac{q_*}{N}\right) \times t_* \tag{1}$$

$$d = \frac{0.5 * \sum_*(Q_{*-1} + Q_*) \times t_*}{q} \tag{2}$$

2.1. The Multigene Genetic Programming model for PPLT

Genetic programming (GP) is a fairly new ML method developed based on Darwin's evolution theory [21]. In contrast to the black-box methods, GP is classified as a grey-box technique with the ability to develop explicit prediction functions. The model transparency is a notable advantage of GP over nearly all other ML methods, especially the "black-box" ANN and DL methods. Searson et al. (2011) introduced a relatively novel formulation of GP called MGGP [20]. In MGGP, a single GP individual is derived from a number of genes where each GP is a tree expression [22]. MGGP combines multiple GP programs via a weighted linear approach.

For the propose of delay modeling in the case of left turn protected/permitted operations, it is important to consider the similar set of variables that the current IQA model utilizes. However, since the goal of the study is to develop a model applicable in real-time operations, an additional parameter called Time-To-Service (TTS) was introduced. TTS serves to indicates when during the r_{pro}, the first vehicle arrived at the intersection stop bar. Such parameter is easily obtainable in most of the intersections utilizing stop-bar detection, and it essentially reveals the amount of red used. To develop a delay model, a function of the following influencing parameters was considered:

$$d = f\left(\frac{AoR_{pro}}{c_{pro}}, \frac{AoG_{pro}}{g_{pro}}, \frac{AoR_{opposing}}{r_{opposing}}, \frac{AoG_{opposing}}{g_{opposing}}, TTS\right) \tag{3}$$

where: d (s/veh) is the average delay per vehicle; AoR_{pro} (*veh*), AoG_{pro} (*veh*), $AoR_{opposing}$ (*veh*), $AoG_{opposing}$ (*veh*), represent the number of vehicles arriving during distinctive periods, c_{pro} (*veh/sec*) is the capacity of a signal group ($c_{pro} =$

$s * g_{pro}/C$; where s (*veh/sec*) represents theoretical maximum flow that can pass through the signal during the unit of time (i.e., 1800 *veh/sec*), C represents cycle length); g_{pro}, $r_{opposing}$, and $g_{opposing}$ represents a duration of particular periods during a cycle as defined previously. Index "opposing" used in previous variables is related to conflicting through movements.

3. Development of delay model using MGGP

3.1. Data collection

For the purpose of this study, we collected the data from a microsimulation model of the real-world arterial located in downtown Chattanooga, Tennessee, US. The performance of exclusive left-turn lanes with protected/permitted operations were observed. Each intersection is equipped with advanced and stop bar detectors that allow the collection of previously defined variables (e.g., AoR, TTS). Data were collected on a cycle-by-cycle manner using PTV Vissim COM and Python scripting [23]. The summary of collected data is presented in Table 1. Overall, data for 140 cycles were obtained and split into three groups consisting of 100, 20, and 20 data instances used for model training, testing, and validation.

Table 1. Descriptive statistics of the variables used for the model development

Parameters	$\frac{AoR_{pro}}{c_{pro}}$	$\frac{AoG_{pro}}{g_{pro}}$ $(\frac{veh}{s})$	$\frac{AoR_{opposing}}{r_{opposing}}$ $(\frac{veh}{s})$	$\frac{AoG_{opposing}}{g_{opposing}}$ $(\frac{veh}{s})$	TTS (s)	Average delay (s)
Mean	0.41	0.011	0.038	0.084	24.5	23.0
Median	0.44	0	0.038	0.071	28.5	22.0
St. Dev.	0.20	0.028	0.021	0.058	11.08	11.1
Minimum	0.22	0	0	0	1	0.8
Maximum	0.88	0.083	0.096	0.214	39	42.8

Assessment of developed models performance was based on several error measures, determination coefficient (R^2), mean absolute percentage error (MAPE), mean squared error (MSE) and mean absolute error (MAE). The models were trained using learning data. Then their performance was checked on validation data. The model with the highest R^2 and lowest RMSE and MAE on learning and testing were selected as optimal.

3.2. Configuration of MGGP and development of delay model

For a number of various parameters, MGGP was utilized to develop an optimum delay model. GPTIPS toolbox, an open resource MGGP training tool coded in MATLAB, was used in this study [20]. A relatively large number for population and generation sizes was used to ensure that the developed model would result in good estimations. In total 3x3x3 = 27 parameter arrangement was used as shown in Table 2.

Table 2. Parameter settings for the MGGP algorithm

Parameter	Setting
Population size	200,400,800
Generations	100, 250, 500
Maximum number of genes	2, 4, 6
Maximum three depth	4
Tournament size	25
Crossover events	0.85
Mutation events	0.1
Function set	+, -, x, /, square

The best performing MGGP-based delay model was obtained and its formulation is given in Equation 4.

$$d = 1.088 * X4 - 1.37 * X3 - 20.26 * X1 + 1.088 * X5 + 38.97 * (X1 - X2)^4 + 20.54 * X1 * X3 - 1.088 * X4 * X5 + \frac{4.478 * 10^{15} * X3 * X4}{2.199 * 10^{12} * X1 + 2.199 * 10^{12} * X5} - X3 * X5 * (X1 - X3)^2 * 6.861 - \frac{1.37 * X2(X1 + X5)}{X1} + 5.664 \tag{4}$$

The tree structure from the proposed MGGP is shown in Figure 2. It should be noted that values from X1, .., X5 stands for input variables ($\frac{AoR_{pro}}{c_{pro}}, \frac{AoG_{pro}}{g_{pro}}, \frac{AoR_{opposing}}{r_{opposing}}, \frac{AoG_{opposing}}{g_{opposing}}, TTS$).

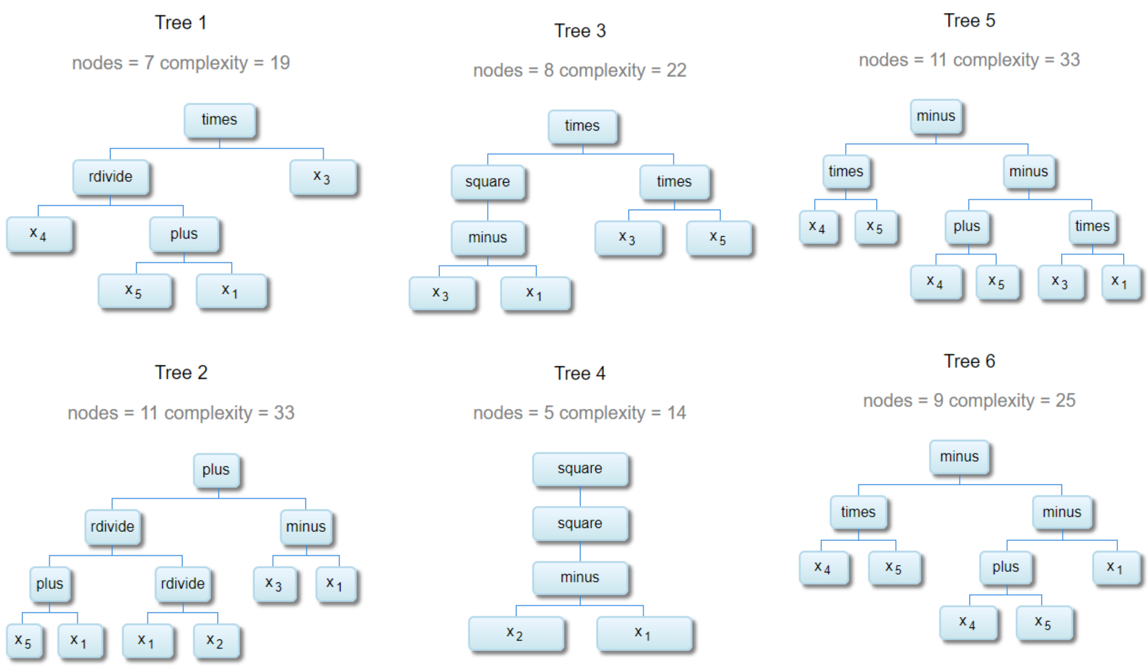

Fig. 2: Expression trees of the best MGGP-based delay model

For the best-performing model, performance measures were calculated and shown in Figure 3. It can be noticed that model achieved satisfactory performance on training, testing, and validation data.

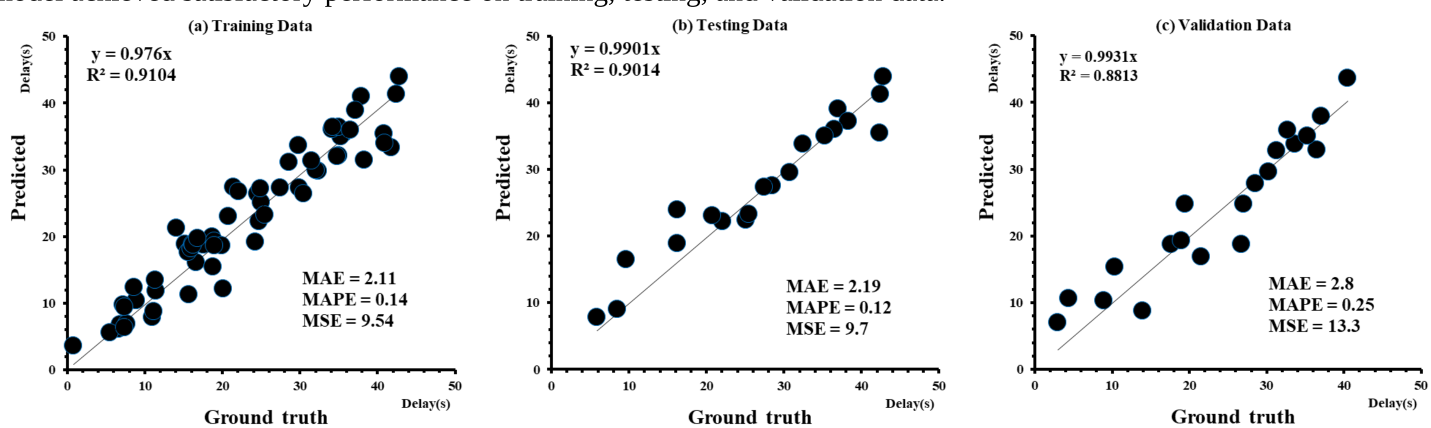

Fig. 3. Statistical values of the GEP prediction models for training (a), testing (b), and validating (c) datasets

3.3. Comparison of IQA and proposed delay model

To compare the developed model with the benchmark IQA model, we exposed both models to a completely new dataset that contains 25 cycles. Figure 4, shows results from this comparison and the accuracy of each model. The proposed model performed significantly better than IQA model, as it can be noticed by observing various performance measures. Due to the complexity of arrival patterns on a cyclical basis, IQA's model assumption provides inaccurate estimates. In some instances, IQA tends to overestimate delays because the model does not account for accurate vehicle arrivals during red, as opposed to the proposed MGGP model (that utilizes the TTS parameter). Also, in many other cases, IQA underestimates delays considering that the model cannot accurately capture delay during a permitted interval caused by platoon-like opposing flow arrivals. Therefore, estimates by IQA are scattered with very low accuracy (see R2 of 0.45). On the contrary, the proposed MGGP model captures these phenomena with acceptable accuracy resulting with R2 = 0.86, MAE = 2.8s, MAPE=0.23 and MSE=12.27. The MGGP model for protected/permitted left turns represents a better alternative for delay estimation for protected/permitted left turns.

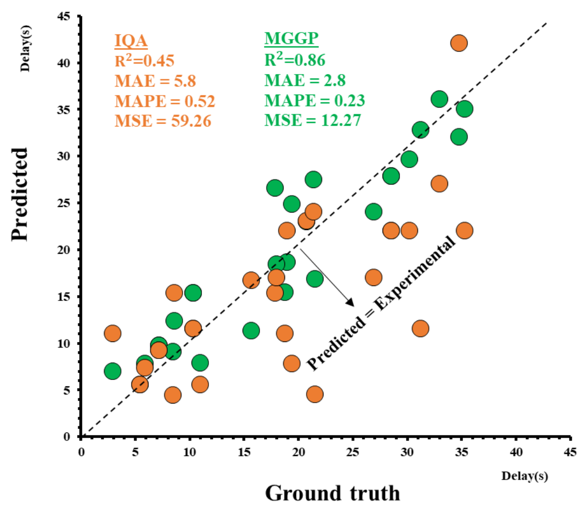

Fig. 4. Comparison of accuracy between IQA and MGGP delay estimates

4. Conclusions

This study utilized Multigene Genetic Programming to model delay for protected/permitted left turns at signalized intersections. The microsimulation model of Martin Luther King Blvd in downtown Chattanooga, Tennessee, US, was used to collect the dataset for the machine learning algorithm. We collected 140 cycles to generate a dataset used for training, testing, and validation purposes. Once the best MGGP model was developed, it was compared with the benchmark IQA model [10] using "unseen" dataset. It was found that the proposed model outperformed the traditional model. In future work, the authors will expand the dataset to cover several hundred cycles and develop a more accurate and reliable delay model to strengthen the conclusions presented here. Also, parameters configuration settings will be investigated more carefully to develop a more robust model that will provide better accuracy. For example, population size and generations should be increased up to 1400, with higher tree depth, and in terms of function set, additional operations should be included (e.g., ^2, ^3, tanh, sin, cos, exp). Future studies will aim to provide more training options for MGGP.

References

[1] Clayton, Arthur John Hamblin. "ROAD TRAFFIC CALCULATIONS." *Journal of the Institution of Civil Engineers* 16.7 (1941): 247-264.

[2] Wardrop, John Glen. "Road paper. some theoretical aspects of road traffic research." *Proceedings of the institution of civil engineers* 1.3 (1952): 325-362.
[3] Webster, F. V. (1958). *Traffic signal settings* (No. 39).
[4] Shayeb, Suhaib Al, Nemanja Dobrota, Aleksandar Stevanovic, and Nikola Mitrovic. "Assessment of Arterial Signal Timings Based on Various Operational Policies and Optimization Tools." *Transportation Research Record* (2021): 03611981211011165.
[5] Nemanja Dobrota, Aleksandar Stevanovic, and Nikola Mitrovic. "Modifying Signal Retiming Procedures and Policies by Utilizing High-Fidelity Modeling with Medium-Resolution Traffic Data." *Transportation Research Record* (2021), 03611981211055668.
[6] Dion, Francois, Hesham Rakha, and Youn-Soo Kang. "Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections." *Transportation Research Part B: Methodological* 38, no. 2 (2004): 99-122.
[7] Little, J. D. (1961). Approximate expected delays for several maneuvers by a driver in Poisson traffic. *Operations Research*, *9*(1), 39-52.
[8] Akcelik, R. (1981). *Traffic signals: capacity and timing analysis*.
[9] HCM. (1985). Highway capacity manual. Washington, DC: Transportation Research Board.
[10] Strong, D. W., Nagui, R. M., & Courage, K. (2006, January). New calculation method for existing and extended HCM delay estimation procedure. In *Proceedings of the 87th Annual Meeting Transportation Research Board*.
[11] HCM. (2016). Highway capacity manual 6th edition, A Guide for Multimodal Mobility Analysis, 2016. *Transportation Research Board.* url: *http://www. trb. org/Main/Blurbs/175169.aspx*.
[12] Kyte, M., Abdel-Rahim, A., Dixon, M., Li, J. M., & Strong, D. (2009). Validating the incremental queue accumulation method for left-turn delay estimation. *Transportation research record*, *2130*(1), 42-51.
[13] Dobrota, N., A. Stevanovic, A., and Mitrovic, N. (2020). Application of Analytical Methods for Traffic Delay Estimation Using High-resolution Signal and Detection Data, Paper presented at the 2021 TRB Annual Meeting, Washington DC.
[14] Manual, Highway Capacity. "Highway capacity manual." Washington, DC 2 (2000): 1.
[15] Wang, Ming-Heng, and Rahim F. Benekohal. "Arrival-based uniform delay model for exclusive protected-permitted left-turn lane at signalized intersections." *Transportation research record* 2027, no. 1 (2007): 91-98.
[16] Yang, Qiaoli, Zhongke Shi, Shaowei Yu, and Jie Zhou. "Analytical evaluation of the use of left-turn phasing for single left-turn lane only." *Transportation Research Part B: Methodological* 111 (2018): 266-303.
[17] Bagdatli, Muhammed Emin Cihangir. "Vehicle delay modeling at signalized intersections with gene-expression programming." *Journal of transportation engineering, Part A: Systems* 146, no. 9 (2020): 04020107.
[18] Murat, Y. Sazi, and Ozgur Baskan. "Modeling vehicle delays at signalized junctions: Artificial neural networks approach." (2006).
[19] Bagdatli, M. E. C., & Dokuz, A. S. (2021). Vehicle Delay Estimation at Signalized Intersections Using Machine Learning Algorithms. *Transportation Research Record*, 03611981211036874.
[20] Searson DP, Leahy DE, Willis MJ (2011) Predicting the toxicity of chemical compounds using GPTIPS: a free genetic programming toolbox for MATLAB. In: Ao S-I, Castillo O, Huang X (eds) Intelligent control and computer engineering, vol 70. Springer, Dordrecht, pp 83–93
[21] Koza, John R., and John R. Koza. *Genetic programming: on the programming of computers by means of natural selection*. Vol. 1. MIT press, 1992.
[22] Gandomi, Amir Hossein, and Amir Hossein Alavi. "A new multi-gene genetic programming approach to nonlinear system modeling. Part I: materials and structural engineering problems." *Neural Computing and Applications* 21, no. 1 (2012): 171-187.
[23] PTV, AG. "PTV Vissim 10 user manual." *PTV AG: Karlsruhe, Germany* (2018)