Traffic Flow Optimization: Modeling the Inter-Arrival Times for Simulation Model

Muhammad Zulqarnain Hakim Abd Jalal[1], Mohd Kamal Mohd Nawawi[2,3], Wan Laailatul Hanim Mat Desa[2], Muhammad Yassar Yusri[1], Nur Fatihah Shaari[4]

[1]Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA Cawangan Johor, Kampus Segamat 85000 Segamat, Johor, Malaysia, [2]School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia, [3]Institute of Strategic Industrial Decision Modelling (ISIDM), Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia, [4]Faculty of Business & Management, Universiti Teknologi MARA Cawangan Johor, Kampus Segamat, 85000 Segamat, Johor, Malaysia
Corresponding Author Email: fatihahshaari@uitm.edu.my

Abstract
With the growing number of vehicles on the road, traffic flow problems are no longer a local issue; instead, traffic flow optimization has drawn significant interest from researchers all over the world. Studies of discrete-event simulation have been widely used to encounter problems related to traffic flow. Researchers of discrete-event simulation modeling typically tend to use statistical distributions for inter-arrival and process times based on the simulation software's built-in tools. The software tools include *Input Analyzer* in *Arena*, *Stat::Fit* for *Promodel*, and *ExpertFit* for *FlexSim.* However, there are other numerical metrics and concerns that researchers should examine while deciding on the best distribution. This research explores the exponential distribution and compares it to distributions generated by software, then to real data. The square error value is the focus of the comparison. There were 5404 data points collected for the vehicle arriving at six lanes at selected traffic junctions. According to the findings of this study, the commonly used exponential distribution can be utilized to depict the distribution of inter-arrival times as there is no significant difference from the more complex distribution. In future study, researcher can comfortably use exponential distribution instead of using complex distribution.
Keywords: Optimization, Discrete-event Simulation, Road Traffic Management, Simulation Modeling, Statistical Distribution

Introduction
With rising consumer awareness of environmental issues, efforts to reduce environmental pollution have received attention from various circles. Hence, it is not surprising that many studies have been conducted to optimize operations in various fields. In this study, the area of focus is on traffic flow. A good management of traffic network might reduce bottlenecks

and ease road traffic congestion. Environmental pollution due to traffic congestion is something that needs to be avoided.

There are many types of modern optimization approaches, but simulation is widely used and seems reliable to cater traffic management problem. Generally, in the traffic engineering and queuing analysis field, the randomness of vehicle time between arrivals is considered exponentially distributed (Adam, 1950; El-Hadidy et al., 2021; Jose, 2021; Kumar, 2022; Meng et al., 2009; Saritha et al., 2022; Sumaryo et al., 2015; Wang et al., 2021). Sumaryo et al (2015) used queuing models M/M/1 and M/G/1 to model traffic flow circumstances. Both models assume that the arrival rate is a Poisson process and that the inter-arrival times are distributed exponentially. This distribution is chosen and assumed based on specific criteria: the number of vehicles on the road is very high, the influence of a single vehicle on system performance is relatively minimal, and all vehicles make their own decisions about whether or not to use the road.

Researchers in discrete-event simulation modelling, on the other hand, tend to use the distributions for inter-arrival and process times recommended by the built-in tools (Frough et al., 2019; El-Hadidy et al., 2021; Jose, 2021; Kumar, 2022; Meng et al., 2009; Saritha et al., 2022; Sumaryo et al., 2015; Wang et al., 2021). Arena Input Analyzer, for example, one of the most widely used tools, provides the best distribution based on the smallest mean square error value. However, Kelton et al (2015) note that there are additional numerical measures and matters that researchers may take into account when deciding on the best distribution. The other two methods for evaluating how well a distribution fits the data are Kolmogorov-Smirnov (K-S) and Chi-square goodness-of-fit tests. The larger the p-value, the better the fit quality is implied by the tests.

Methodology
Data Collection
The data for this study came from a large public traffic light junction on the AH2 highway, which connects Malaysia and Thailand. In this paper, we use the input data from the (Jalal et al., 2017 Jalal et al., 2018). Data was gathered through observation and interviews. Traffic light junction operation is a continuous process with variable traffic flow. Traffic may be congested for some time due to the large number of vehicles on the road, but it may be smooth for other parts of the day.

Several observations were made to obtain information on four types of data: traffic light control (TLC) cycle pattern, TLC cycle time, vehicle arrival at each junction, and model conceptualization. In addition, interviews with various teams involved in TLC operations have taken place. A few Public Works Department (PWD) employees shared important information about TLC operations, which helped to improve understanding of TLC operations. An interview was also conducted with road users who are frequently involved in road traffic congestion (RTC) on the Changloon main road. From there, useful information was gathered, which formed a solid foundation for this study. Figure 1 shows the snapshot of Malaysia border area with Thailand at Kedah while Figure 2 depicts the main road layout Changloon.
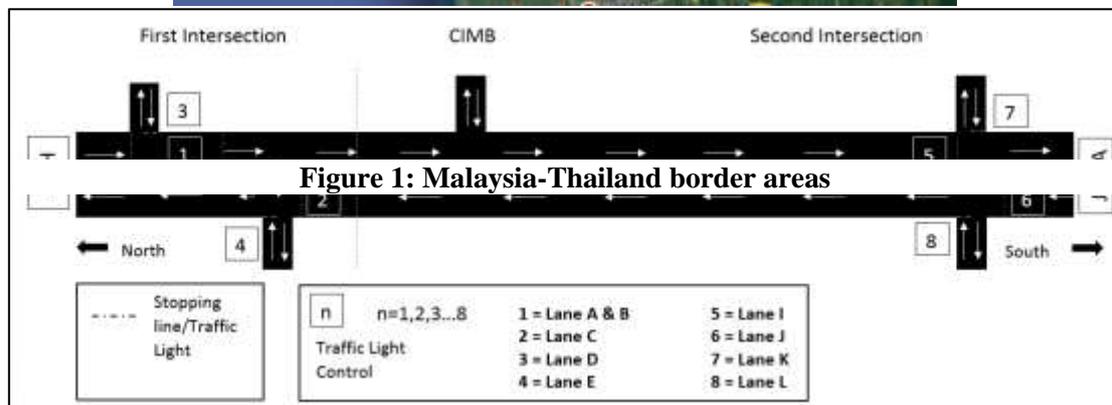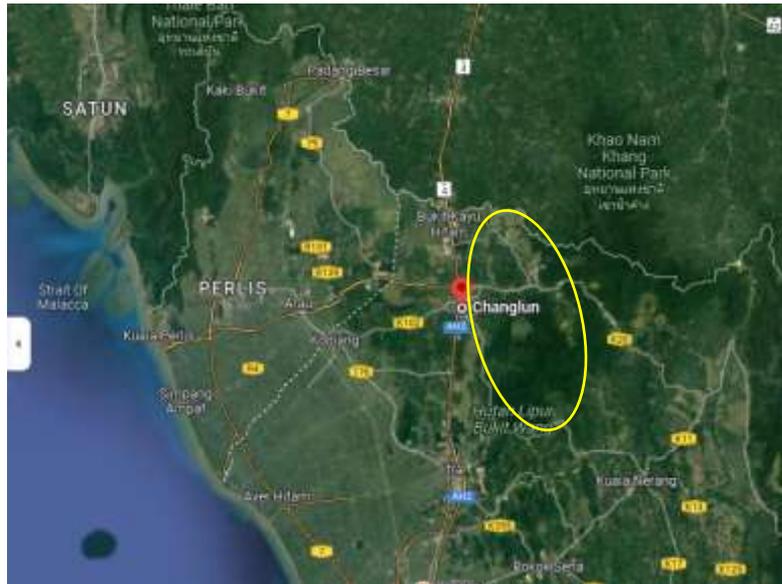
Figure 1: Malaysia-Thailand border areas

Figure 2: Hierarchy system of the liner selection

Input Modeling

Specific input information, such as the arrival rate, is needed when creating a simulation model in Arena. Input Analyzer, an Arena built-in feature, was used to gather and analyse the inter-arrival times between vehicles for this study. A probabilistic expression is given by the input analyzer and used to build the model. Figure 3 displays the results of an exponential and suggested distribution for the inter-arrival time at junctions for the first intersection, Lane A. The software recommended the lognormal distribution because its square error value, 0.002434, is lower than that of other distributions.
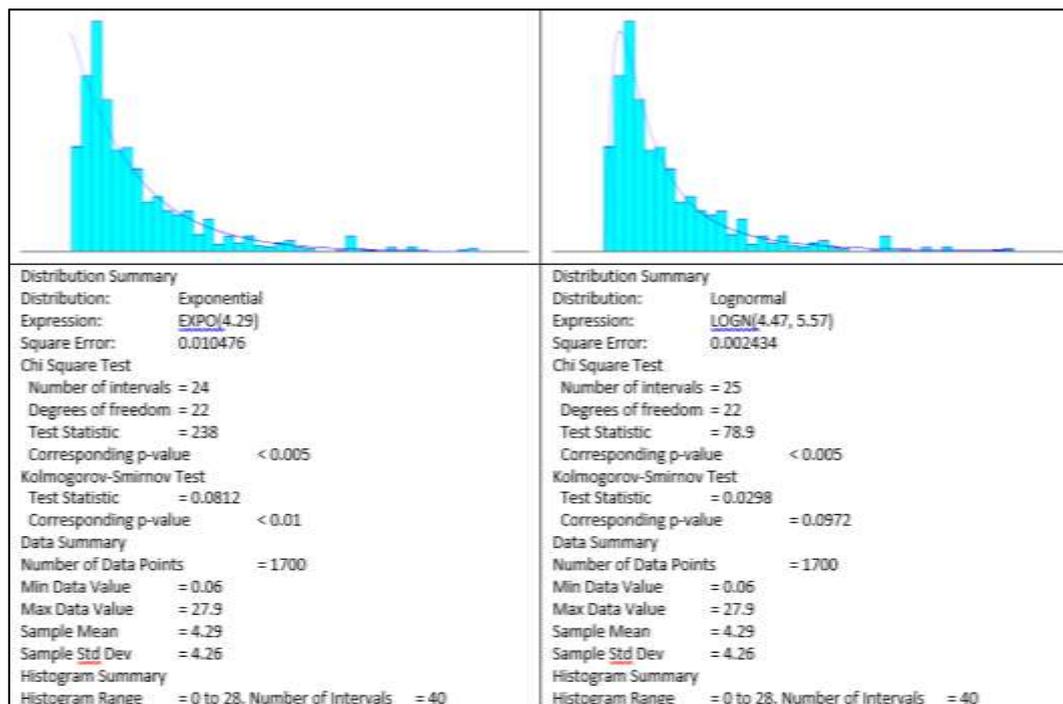
Figure 3: Exponential and suggested distributions for Lane A.

In contrast, the square error value for an exponential distribution is 0.010476. With a mean of 4.29, the suggested exponential distribution for Lane D suggests that this distribution is common and frequently used for inter-arrival time. The results of the Lane D inter-arrival time are shown in Figure 4.
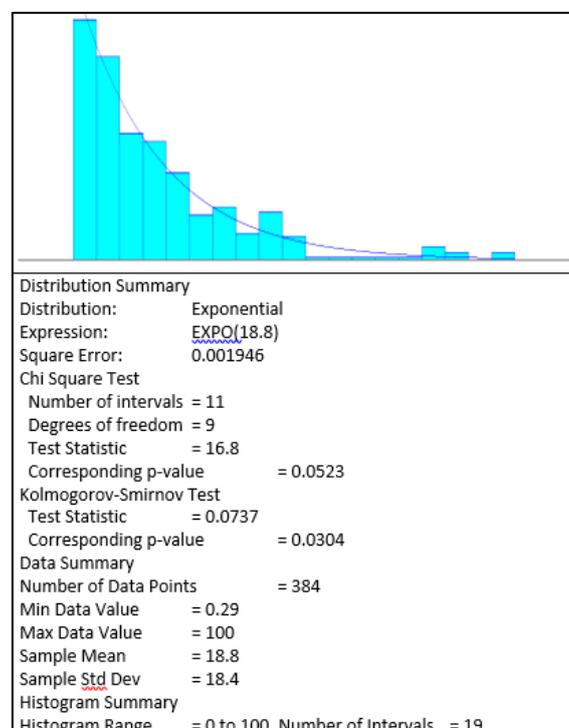

Figure 4: Suggested distributions for Lane D.

The output for the inter-arrival time at the junction for Lane E using the exponential and suggested distribution is shown in Figure 5. The software recommended the Weibull

distribution because its square error value, 0.002611, is lower than that of other distributions. In contrast, the square error value for an exponential distribution is 0.003752.

Figure 5: Exponential and suggested distributions for Lane E.

Figure 6 depicts the output obtained using an exponential and suggested distribution for the inter-arrival time at the Lane J junction. The software recommended the lognormal distribution because its square error value, 0.000944, is lower than that of other distributions. In contrast, the square error for an exponential distribution is equal to 0.010112.
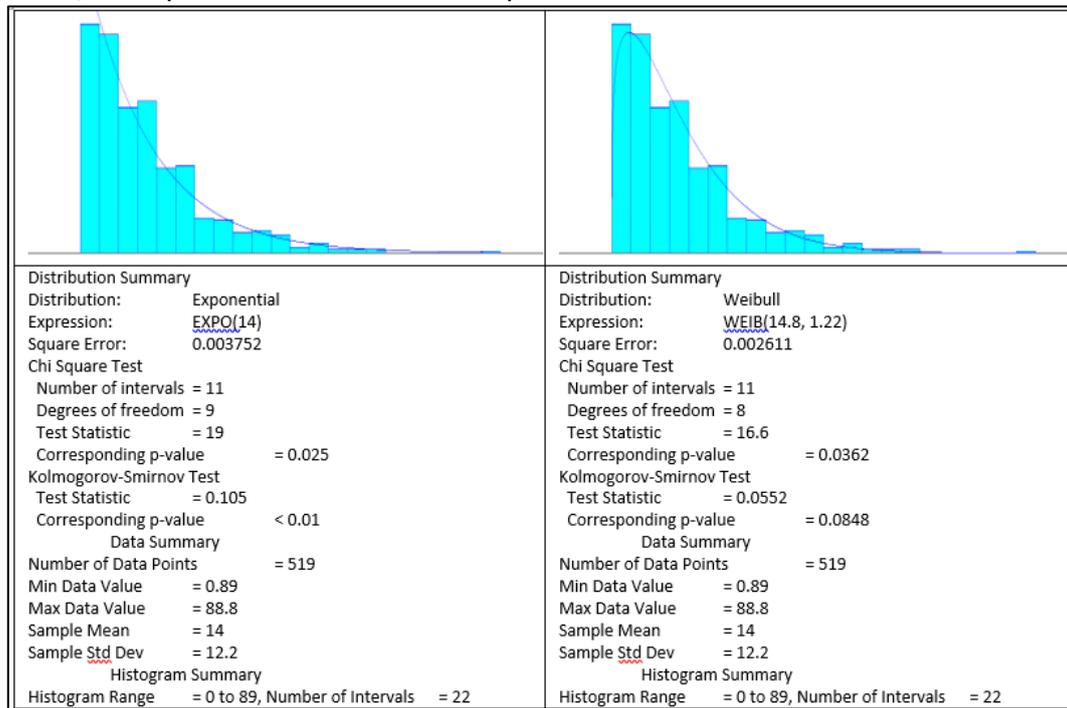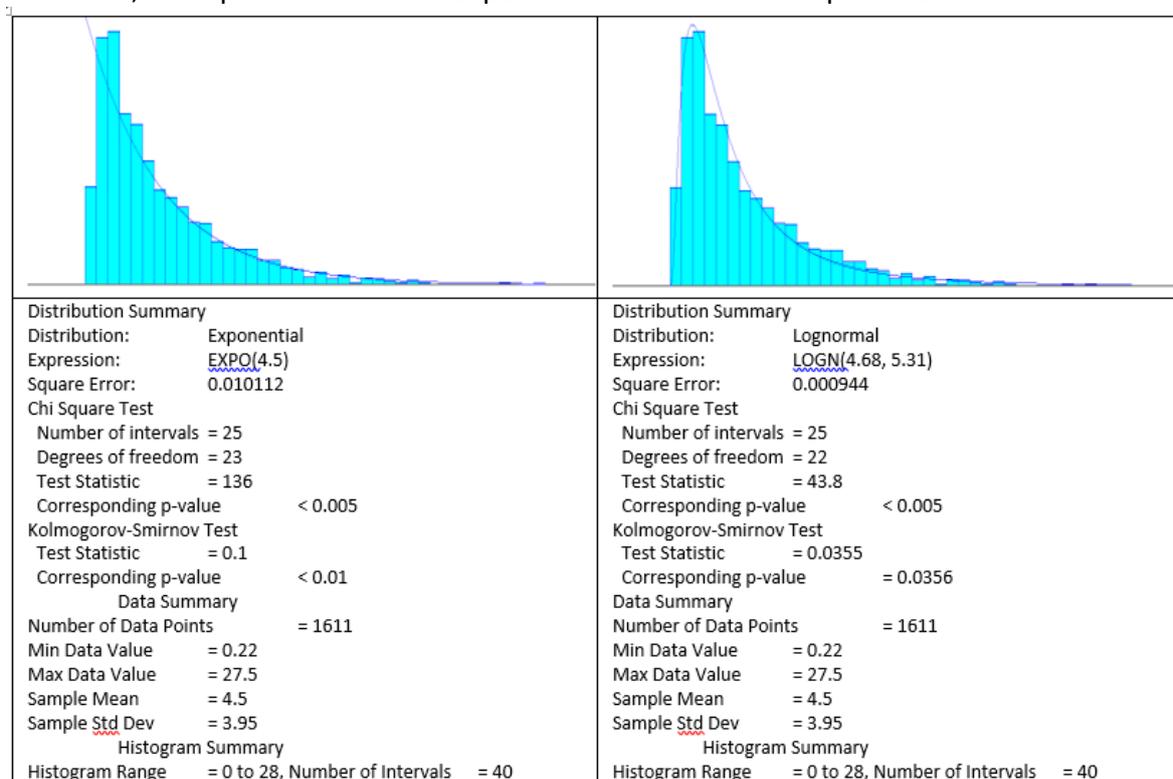

Figure 6: Exponential and suggested distributions for Lane J.

The proposed distribution for Lane K is exponential, with a mean of 10.6, indicating that exponential is a common and widely used distribution for inter-arrival time. The output of Lane K inter-arrival time is shown in Figure 7.

Finally, using an exponential and suggested distribution, Figure 8 displays the output for inter-arrival time for Lane L. The software recommended this kind of distribution because the square error value for a lognormal distribution, 0. 001481, is smaller than that for other distributions. In contrast, the square error value for an exponential distribution is 0. 003207.


```
Distribution Summary
Distribution:         Exponential
Expression:          EXPO(10.6)
Square Error:         0.000003
Chi Square Test
 Number of intervals  = 1
 Degrees of freedom   = -1
 Test Statistic        = 0.000589
 Corresponding p-value          < 0.005
Kolmogorov-Smirnov Test
 Test Statistic        = 0.173
 Corresponding p-value          < 0.01
Data Summary
Number of Data Points           = 680
Min Data Value        = 0.86
Max Data Value        = 2.07e+003
Sample Mean           = 10.6
Sample Std Dev        = 79.5
Histogram Summary
Histogram Range       = 0 to 2.07e+003, Number of Intervals     = 26
```

Figure 7: Suggested distributions for Lane K.

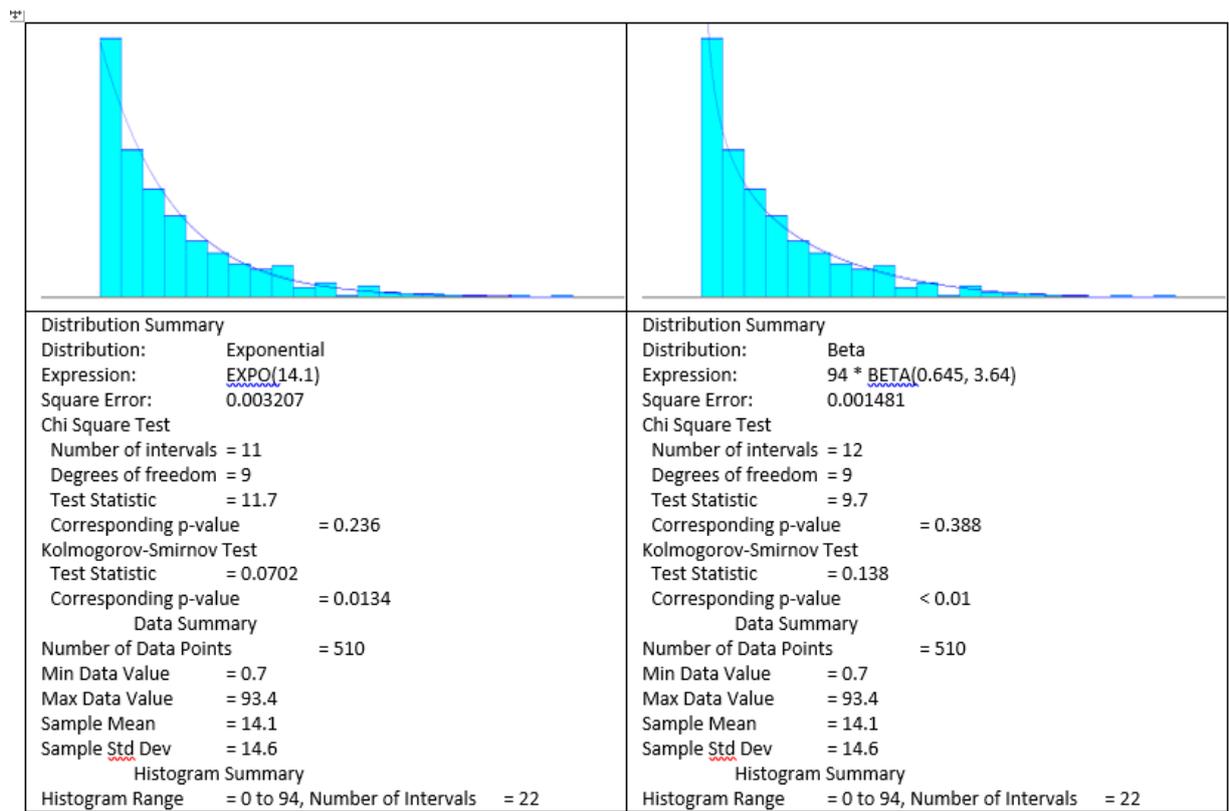| Distribution Summary		Distribution Summary	
Distribution:	Exponential	Distribution:	Beta
Expression:	EXPO(14.1)	Expression:	94 * BETA(0.645, 3.64)
Square Error:	0.003207	Square Error:	0.001481
Chi Square Test		Chi Square Test	
Number of intervals = 11		Number of intervals = 12	
Degrees of freedom = 9		Degrees of freedom = 9	
Test Statistic = 11.7		Test Statistic = 9.7	
Corresponding p-value	= 0.236	Corresponding p-value	= 0.388
Kolmogorov-Smirnov Test		Kolmogorov-Smirnov Test	
Test Statistic = 0.0702		Test Statistic = 0.138	
Corresponding p-value	= 0.0134	Corresponding p-value	< 0.01
Data Summary		Data Summary	
Number of Data Points	= 510	Number of Data Points	= 510
Min Data Value = 0.7		Min Data Value = 0.7	
Max Data Value = 93.4		Max Data Value = 93.4	
Sample Mean = 14.1		Sample Mean = 14.1	
Sample Std Dev = 14.6		Sample Std Dev = 14.6	
Histogram Summary		Histogram Summary	
Histogram Range = 0 to 94, Number of Intervals = 22		Histogram Range = 0 to 94, Number of Intervals = 22	

Figure 8: Exponential and suggested distributions for Lane L.

Model Development

Arena simulation software was used to create the model. The distributions of the inter-arrival time, rather than the service or process time, are the focus of this study. The process times for all traffic lights at the junctions are all deterministic and do not require any input distribution analysis. The process times or durations of the green lights are 120 seconds (Lane A), 30 seconds (Lane D), 35 seconds (Lane E), 60 seconds (Lane J), 50 seconds (Lane K), and 45 seconds (Lane K) (Lane L).

Figure 9 depicts an Arena modelling layout at the Lane L junction. The modelling process began with a Create module that represented the vehicle's arrival at the intersection. The vehicle type was assigned using an Assign module. The entity was then moved to a decide module, and eighty percent of the entities were distributed to hold module "Lane L," while the remaining percentage went to enter junction two. The Hold module waits for the departure signal, and only one entity is permitted to depart once the signal is received. The entity was then moved to a decide module and distributed to the exit points as needed. Jalal et al. (2017) provides a detailed explanation of model development.
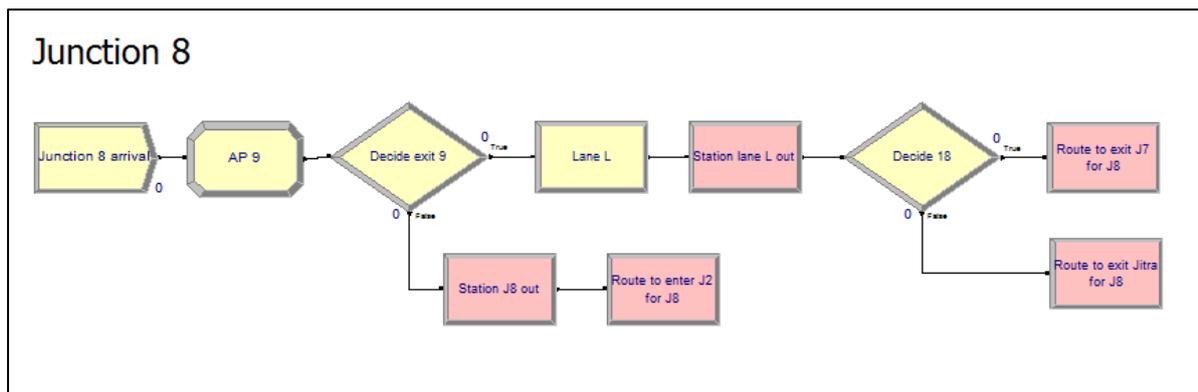
Figure 9: Example of Arena flowchart module for Lane L.

Results
The distributions are generated from thousands of input data points collected from two traffic light junctions in Changloon. Saritha et al. (2022) contains a detailed description of the junctions.

Input Data
There were 5404 data points collected for vehicles arriving between 5.00 p.m and 7.00 p.m. at all six junctions. As shown in Table 1, these data can be further subdivided into each junction.

Table 1
Data collected for the number of vehicle arrival

Lane	Data (t)	Distribution A	Distribution B
Lane A	1700	EXPO(4.29)	LOGN (4.47, 5.57)
Lane D	384	EXPO (18.8)	EXPO (18.8)
Lane E	519	EXPO(14)	WEIB (14.8, 1.22)
Lane J	1611	EXPO(4.5)	LOGN (4.68, 5.31)
Lane K	680	EXPO (10.6)	EXPO (10.6)
Lane L	510	EXPO(14.1)	94 * BETA(0.645, 3.64)
Total	**5404**		

The arrival rate for vehicles is commonly assumed to be exponentially distributed in traffic engineering and queuing analysis. Comparisons were made in this study using six data sets, one for each lane.

Model Validity
Building the right model is a key component of model validation. It's crucial to stress that no model is ever fully validated or verified. Any simulation model is a simplified representation of a real system, and the behaviour of the model is only a rough approximation of the behaviour of the system (Carson, 2002; Desa et al. 2015). A model is said to have been verified or validated when the modeller has explicitly completed a number of tasks to verify and validate the model to the extent required for his or her purposes. Such approval is almost always largely a matter of judgement. Model validation and verification increase the decision maker's confidence in the model, which is referred to as model credibility.

The output for particular parameters was taken into consideration, and it was thought to be compared with the actual value from the data gathered. We applied the validity level used in earlier studies by Hashim et al. (2003) in this study, which stated that the output should be within 10% of actual data. The following formula can be used to determine the validity level:

$$\frac{[Output\ (data) - Output\ (simulated)]}{Output\ (data)} \text{ x } 100\% \leq 10\%$$

As shown in Table 2, the validity levels for all lanes were lower than 10% which indicates that both exponential and software suggested distributions are valid to be used in the model. Even, with exponential distribution, simulated results give lower differences or better validity levels for Lanes A and J. For Lanes D and K, the results are the same since the software suggested exponential as the best distribution

Table 2
Data collected for the number of vehicle arrival

Lane	Actual data	Simulated A	Validity level A	Simulated B	Validity level B
A	1700	1670.48	1.74%	1609.55	5.32%
D	384	387.79	0.99%	387.79	0.99%
E	519	516.20	0.54%	517.82	0.23%
J	1611	1608.94	0.13%	1542.10	4.28%
K	680	683.85	0.57%	683.85	0.57%
L	510	513.13	0.61%	510.49	0.10%

Contribution of Study
Developing a simulation model is a process to imitate the real-world situations. A good simulation model is known to have a small difference between simulated value and actual value, which lead to high validity level. A built-in feature of Arena software called Input Analyzer is helpful in the effort to create a good distribution for arrival time because it can provide the best fit distribution and is not just limited to exponential distributions. From the findings, either exponential or not, the results have no significant differences. Instead of using a complex distribution, the developer can easily design the model using a simpler exponential distribution.

The outcomes are expected to provide the guidelines and perhaps, able to give awareness to the respective authorities which are responsible for handling the simulation study. It is hoped that with the proposed solutions, many parties can benefit.

Conclusion
This paper compares several statistical distributions based on real-world input data of time between arrivals or inter-arrival time. There are no significant differences between the suggested distribution and the findings, according to the results. In this study, we conclude that a simple and well-known distribution, such as the exponential distribution, can be used to build a valid simulation model instead of other complicated distributions.

References

Adams, W. F. (1950) "Road Traffic Considered as a Random Series," *Oper. Res. Q.*, vol. 1, no. 1, p. 9, 1950.

Desa, W. L. H. M., Kamaruddin, S., Nawawi, M. K. M., and Zulkepli, J (2015) "Evaluation on absenteeism effect in production line at aircraft composite manufacturer," *J. Teknol.*, vol. 77, no. 5, pp. 63–67, 2015, doi: 10.11113/jt.v77.6120.

El-Hadidy, M. A. A., and Alfreedi, A. A. (2021) "Detection of an appropriate pharmaceutical company to get a suitable vaccine against COVID-19 with minimum cost under the quality control process," *Qual. Reliab. Eng. Int.*, vol. 37, no. 6, pp. 2646–2664, 2021, doi: 10.1002/qre.2881.

Frough, O., Khetwal, A., and Rostami, J. (2019) "Predicting TBM utilization factor using discrete event simulation models," *Tunn. Undergr. Sp. Technol.*, vol. 87, pp. 91–99, 2019, doi: 10.1016/j.tust.2019.01.017.

Hashim, S., Tahar R. M., Bakar E. M. N. E. A. (2003) "Simulation study for improving patient treatment services," *J. ICT*, 2003.

Jalal, M. Z. H. A., Nawawi, M. K. M., Desa, W. L. H. M., Khalid, R., Abduljabbar, W. K., and Ramli, R. (2017) "Green supply chain: Simulating road traffic congestion," in *Journal of Physics: Conference Series*, 2017, vol. 890, no. 1, doi: 10.1088/1742-6596/890/1/012111.

Jalal, M. Z. H. A., Desa, W. L. H. M., Nawawi, M. K. M., Khalid, R. (2018) "Discrete-event simulation of road traffic congestion to support green supply chain," *Int. J. Eng. Technol.*, vol. 7, no. 3.20 Speci, pp. 377–380, 2018.

Jose, J. K., and Deepthi, V. (2021) "M/PH/1 queueing model with re-servicing," *Commun. Stat. Simul. Comput.*, 2021, doi: 10.1080/03610918.2021.1921209.

Kelton, D. W., Sadowski, R. P., and Zupick, N. B. (2015) *Simulation with Arena*, 6th ed. New York: McGraw-Hill, 2015.

Kumar, B. (2022) "Transient analysis of M/M/C queuing model with reneging, finite capacity and population," *Int. J. Math. Oper. Res.*, vol. 21, no. 1, pp. 1–25, 2022, doi: 10.1504/ijmor.2022.120339.

Meng, Q., and Khoo, H. L. (2009) "Self-similar characteristics of vehicle arrival pattern on highways," *J. Transp. Eng.*, 2009, doi: 10.1061/(ASCE)0733-947X(2009)135:11(864).

Neeraj, R. R., Nithin, R. P., Niranjhan, P., Sumesh, A., and Thenarasu, M. (2018) "Modelling and simulation of discrete manufacturing industry," in *Materials Today: Proceedings*, 2018, pp. 24971–24983, doi: 10.1016/j.matpr.2018.10.298.

Saritha, S., Mamatha, E., Reddy, C. S., and Rajadurai, P. (2022) "A model for overflow queuing network with two-station heterogeneous system," *Int. J. Process Manag. Benchmarking*, vol. 12, no. 2, pp. 147–158, 2022, doi: 10.1504/IJPMB.2022.121592.

Sumaryo, S., Halim, A., and Ramli, K. (2015) "Simulation and analysis of traffic flow models with emergency vehicles distortion on a single road," 2015, doi: 10.1109/TIME-E.2014.7011611.

Wang, C., Chen, W., Xu, Y., and Ye, Z. (2021) *Modeling bus capacity for bus stops using queuing theory and diffusion approximation*, vol. 2675, no. 12. 2021.

Yemane, A., Gebremicheal, G., Meraha, T., and Hailemicheal, M. (2020) "Productivity improvement through line balancing by using simulation modeling (case study almeda garment factory)," *J. Optim. Ind. Eng.*, vol. 13, no. 1, pp. 153–165, 2020, doi: 10.22094/JOIE.2019.567816.1565.

Zulkepli, J., Khalid, R., Nawawi, M. K. M., and Hamid, M. H. (2017) "Developing a discrete event simulation model for university student shuttle buses," in *AIP Conference Proceedings*, 2017, vol. 1905, doi: 10.1063/1.5012146.