World Academy of Science, Engineering and Technology
International Journal of Civil and Environmental Engineering
Vol:2, No:12, 2008

TOWARDS MODELING FOR CRASHES
A Low-Cost Adaptive Methodology for Karachi

Mohammad Ahmed Rehmatullah

Abstract**—** The aim of this paper is to discuss a low-cost methodology that can predict traffic flow conflicts and quantitatively rank crash expectancies (based on relative probability) for various traffic facilities. This paper focuses on the application of statistical distributions to model traffic flow and Monte Carlo techniques to simulate traffic and discusses how to create a tool in order to predict the possibility of a traffic crash. A low-cost data collection methodology has been discussed for the heterogeneous traffic flow that exists and a GIS platform has been proposed to thematically represent traffic flow from simulations and the probability of a crash. Furthermore, discussions have been made to reflect the dynamism of the model in reference to its adaptability, adequacy, economy, and efficiency to ensure adoption.

Keywords**—** Heterogeneous traffic data collection, Monte Carlo Simulation, Traffic Flow Modeling, GIS.

I. INTRODUCTION

KARACHI is growing at a phenomenal rate. With past incentives towards car leasing and low import duties on cars on one hand, and the deteriorating public transportation on the other; the city faces an unprecedented growth in vehicle ownership. This in turn is resulting in a massive decline in traffic flow conditions all round the city and thus traffic jams, congested flow, and traffic accidents are a certainty Karachites have to face every day. According to Helman, as much as 40% road traffic congestion occurs due to traffic incidents, work zones and poor signal timings [1]. Therefore, catering for crashes not only improves traffic safety, but also alleviates congestion issues.

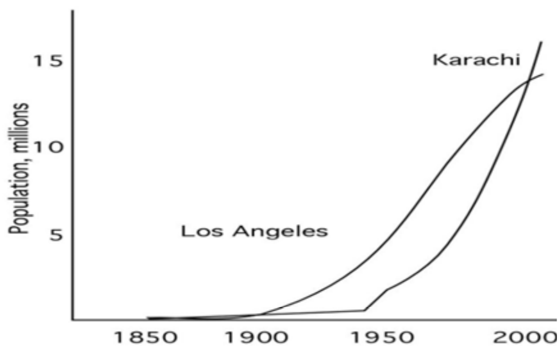

Fig. 1 Growth rate of Karachi compared to that of Los Angeles [2]

Mohammad Ahmed Rehmatullah is a Lecturer at the Department of Civil Engineering, NED University of Engineering and Technology, Karachi 75260, Pakistan. (Phone: (9221) 9261261-68; fax: (9221) 9261255; e-mail: marehmat@ neduet.edu.pk).

It is safe to assume that there is a relationship between traffic flow and crashes. Also, the three interrelated traffic flow characteristics, namely speed, flow and density, all have pronounced influence on traffic accidents [3]. So, modeling traffic flow parameters could in turn help us quantify the expected (or probabilistic) number of crashes. This paper proposes a statistical approach toward traffic flow modeling. Alongside, Monte Carlo Simulation techniques have been adopted in this paper for the simulation of traffic flow models. Thematic representation of simulation results is represented on a GIS (Geographic Information Systems) platform, which makes visualization of crash predictions a lot easier.

Furthermore, a cost effective methodology is devised to ensure implementation. The methodology is also meant to be efficient and adaptive so that it may cater for further advancements and upgrades in data collection and analysis techniques.

II. METHODOLOGY

The general flow of work is depicted in Fig. 2. The process

Fig. 2 Process Flow diagram for Crash Prediction

begins with extensive data collection due to unavailability of appropriate traffic data. The data collection procedures accommodates for the heterogeneous, undisciplined lane-changing behavior of the traffic flow here in Karachi. After data collection, data extraction and initial analysis, a statistical traffic flow model is to be calibrated to fit. Pearson type III distribution is used to model traffic, as this type incorporates a family of distributions including negative exponential distribution, Erlang distribution and normal distribution [4]. Once a flow model is developed, a random traffic data set, for the specific epoch, would be used to test the model and verified. Monte Carlo Simulation techniques would then be employed to simulated traffic and calculate the probability of conflict. The traffic flow simulation outputs would be compared to real flow to assign a level of confidence to our predicted values. Simulated traffic flows would then be analyzed for the calculation of the probability of crash. Finally, the simulated traffic flow and the probability of crashes would be loaded onto a GIS platform of the transportation facility under analysis for thematic representation.

III. DATA COLLECTION

Traffic flow model is one of the most integral part of this project. Data collection for traffic flow modeling, therefore, is critical. Microscopic flow is modeled by the use of time headways. The problem, however, lies elsewhere. Most of the flow models are developed for homogeneous traffic flow where passenger cars dominate the traffic flow. Such models cannot be directly incorporated for cities such as Karachi, as most major cities in the developing countries have vehicles with diverse physical and dynamic characteristics [5]. Random and frequent lane change behavior and indifference to traffic lanes are other characteristics linked with road traffic flow in Karachi. Therefore, a new data collection methodology was developed, where heterogeneous traffic, with the incorporation of a few extra parameters during data collection, may be approximated to flow as homogenous traffic. This can be achieved via a three step process.

A. *Step One*

Vehicular time headways are measured along with their classifications as if they followed lane behavior i.e. all vehicular headways are measured that completely or partially lie in the specific lane under observation. At places where lane markings are present, the methodology is straight forward. Roadways where lane markings are absent, a 10 feet lane width was assumed.

B. *Step Two*

A binary ‘nuisance factor’ is associated with every vehicle which represents the level of presence each vehicle has on that specific lane. This factor is binary as the vehicle may or may not influence the flow of the vehicle following. This factor, thus, accommodates for the lack of lane discipline as the partial presence of a vehicle on a lane may affect the flow in that lane. When this is the case, the factor is taken as 1 and the vehicle is considered to be flowing in the specified lane. If the partial presence does not hinder flow, it is accounted for as zero and the vehicle is assumed not to be a part of the specified lane. Several tests and trial runs made has validated this data collection methodology.

C. *Step Three*

Keeping in line with the collection methodology in steps one and two, step three required us to measure the time taken by a vehicle (along with its vehicular class) to clear a speed trap. A measured road length acts as a speed trap and the arrival time and departure times are noted per vehicle. Spot speed can hence be calculated for a vehicle of a specific class in a defined epoch.

Due to inadequacy of existing traffic data, data collection is an integral part of this exercise. As quite a few parameters are to be observed per vehicle, manual data collection was concluded inefficient. Also, the data collection methodology needed to be cost effective, efficient and is not alter the vehicular flow behavior i.e. cheap non-intrusive methodologies need to be applied. A simple solution was devised to cater for these issues. A camcorder was used to shoot videos at data collection sites which were then analyzed on a computer later. 16 hours of video survey was performed per day for four days per site to have a representative sample size. The camera was mounted on a heavy duty tripod at locations and angles to ensure optimal view of the road length under study. Reference markings and measurements were also taken during the survey for speed calculations later. These videos were then imported to a computer and analyzed for time headway and speed. An on-computer stopwatch was used in order to measure time headways. Headway measurements are made for almost every vehicle regardless of their classification. Speed measurements were made using the distances from reference markings made earlier and the time taken to clear the speed trap. Here the classification was considered important. Measurement errors of distance and time were also incorporated to improve data quality. [6] After data extraction from videos, the headway and speed data were grouped together. Some logical and statistical checks were applied to remove erroneous data and outliers. Time headways were grouped in intervals of 0.5 seconds, where as spot speeds were grouped in intervals of 2 km/hr for respective traffic classifications.

IV. TRAFFIC MODELING AND VERIFICATION

A. *Modeling for Flow*

The time headway between vehicles is an important flow characteristic that affects safety. The distribution of time headway determines the requirements and the opportunity for passing, merging and crossing. [7] Statistical distribution models also help in making predictions for conflicts with minimal amount of information. Alongside they are useful in describing phenomena with high degrees of randomness and therefore statistical distributions have been used to model traffic flow effectively.

Traffic flow can be categorized to be either in random, intermediate, or uniform state. Modeling for every state requires a different distribution namely the negative exponential distribution for traffic flow with random headway, Pearson type III distribution for traffic flow with intermediate headway and normal distribution for traffic flow with uniform headway. Incorporation of all the above mentioned distributions towards modeling traffic flow for simulation is complex and tedious. Luckily the Pearson type III distribution accommodates both negative and normal distributions as well, thus simplifying the problem.

The Pearson type III distribution, a part of the twelve Pearson distributions, is of particular use as it accommodates negative exponential distribution, Erlang distribution, gamma distribution, and normal distribution along with their shifted counterparts. The probability density function for Pearson type III distribution is given by (1)

$$f(t) = \frac{\lambda}{\Gamma(K)}\left[\lambda(t-\alpha)\right]^{K-1} e^{-\lambda(t-\alpha)} \quad (1)$$

where K is the shape parameter, α is the location parameter, λ is the shape parameter, and t is the time headway.

The traffic data extracted from video feeds was analyzed for 15 minute time segments, grouped in class width of 0.5 seconds (time headway). After the data from the sample is ready, the values of K, α, and λ are deduced. Once done, we get the vehicular flow model for that specific time frame and lane.

It is the values of K, α, and λ which dictate the type of headway distribution being followed. This can be represented graphically as Fig. 3.

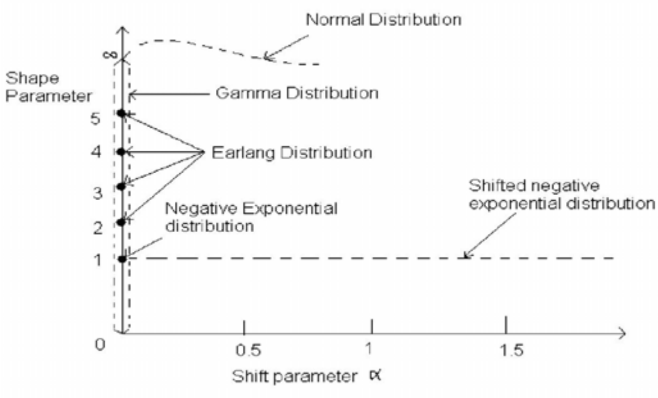

Fig. 3 Pearson type III distribution, values of K and **α**. [7]

The values of K, α, and λ of the traffic flow was deduced by using the detailed stepwise description presented by May (1990) [7]. However an alteration was made to this methodology. It was found that the α value, which may range from 0 to infinity inclusive, usually ranged between 0 and the average time headway. This limitation was successfully adopted towards modeling of flow.

The model is, however, only valid for simulation purposes with the "best" Pearson type III distribution parameters [7]. Therefore, in order to identify the "best" values, an iterative process was adopted. With the use of the VBA (Visual Basic for Applications) platform and macro-enabled spreadsheets, a range of α and K values (and therefore λ values) are sequentially placed in the distribution equation and tested. The resultant equation was then tabulated and compared with the actual traffic via extracted data from traffic videos. The Kolmogrov-Smirnov (KS) goodness-of-fit test was used to test the differences between the model and the field data. KS test is used specifically ahead of Chi-square test because Chi-square test has the ability to invalidate the test because of small variance in data. On the contrary, the KS test produces statically good fits to the data. KS tests is also has simpler calculations and thus is more powerful than Chi-square test. [8]

As the K can further be estimated using the value of α, the macro was designed such that for a specific range of values for α, a specified range of K was generated. Therefore, each combination of α and K values were inserted into the equation, tabulated, and the KS test was run. This is done for the entire range of the estimated α and K values and the best fit values selected.

This process results in a traffic flow prediction equation. As this equation formed from the sample data of a population, the model is verified using a secondary sample of the same epoch as that for the model. The results showed that the model and the sample showed statistical homogeneity, thus verifying the model's result.

B. Modeling for Vehicular Speed and Classification

Various vehicle classes travel within various speed ranges within a traffic stream. These speed ranges may vary greatly between the classes at different levels of service. This requires us to take a two staged approach towards modeling for speed and classification. Primarily, identification of different classes is critical prior to the data collection phase. Stage One involves enlisting each class of vehicle in the traffic flow during the epochs of analysis and collecting data for spot speed calculation individually for each vehicle class. At stage two, vehicular speed of individual classification is to be modeled. This can be done using the same principles as used in modeling flow parameters done previously. Here, Pearson type III distribution can be readily deployed as it can cater for normal and skewed normal distributions.

The modeling exercise is performed to deduce the distribution of speed along its vehicular class only. Vehicular classification can be safely assumed to be following uniform distribution were the probability of occurrence of any vehicle is unbiased and dependent only upon its percent share of traffic volume.

V. SIMULATION

The study of complex systems that cannot be sufficiently simplified to be amenable to analytical solution requires alternative methods; the use of simulation is one possibility. [9] One of the greatest advantages of simulation, as opposed to observation of actual traffic streams, is the ability to control all the conditions related to the traffic streams. [10] Once the results from the modeling exercise are verified, they are used to simulate traffic. With the shape factor, the shift factor, and size factor in place, the Pearson type III distribution may be used in the Monte Carlo Simulation method. This technique is just one of many methods for analyzing uncertainty propagation, and can be any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property [11], which, in our case, is the probability of crash. The simulation is performed in three tiers namely the simulation of traffic flow, vehicle classification, and finally, vehicle speed.

A. Tier One: Simulation of Traffic flow

The simulation of traffic flow is dependent on the time headway distribution. With the determination of the internal parameters of the Pearson type III distribution, our traffic flow model for simulation is complete. Using the principles of Monte Carlo method, random numbers are generated on a spreadsheet which acts as the probabilities for the cumulative probability function of the calibrated Pearson type III distribution. This results in the expression to become a univariate function, where the variable is the time headway. The successive results to the equation for unbiased random numbers simulate headways to following vehicles, thus simulating traffic flow.

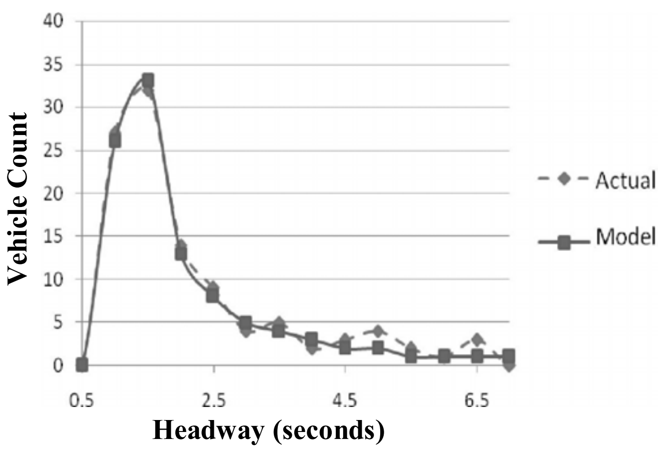

Fig. 4 Graphical representation of simulated traffic flow against actual flow at a major road in Karachi during off-peak hours.

B. Tiers Two and Three: Simulation of Vehicle Classification and Speed

Very similar to tier one, tiers two and three simulate classification and speed of the flowing vehicles. Primarily, simulation of vehicle classification is done with the use of uniform distribution. The uniform distribution has been adopted so that each class of vehicle has a probability of occurrence equal to its share in the total flow volume for that specific time segment. Once the vehicle class has been simulated, the appropriate speed distribution model for that specific class, calibrated previously, can be adopted for simulation.

Monte Carlo method was used to simulate these parameters for flow in correspondence with the flow simulation in tier one.

It is to be noted that different random numbers were generated for all the individual tiers to prevent bias simulation results. Once simulation is complete, the output results are tabulated such that each vehicle simulated is marked with its time headway, a vehicle class and a speed. Alongside, the occurrence of each vehicle scale is tabulated on an absolute time, depicting the time of occurrence for the vehicle once the simulation has started.

VI. CONFLICTS AND PROBABILITY OF CRASH

The probability of crash can be determined by combined analysis of multiple flows from various directions. The three conflict types, the crossing, the converging, and the diverging, needs to be realized forehand. This methodology focuses on quantifying the probability of conflict rather than the intensity of a crash.

With the help of the simulation model, traffic flow can be generated with varying headways, speeds, and classifications. To calculate the probability of crash, simulations of two different traffic flow streams are compared. The prediction of probabilities is done in three steps.

A. Step One

Calculate the total stopping time (say ‘p’ seconds) for a vehicle traveling at average speed of the flow.

B. Step Two

Identify the vehicles which would occur at the point of conflict within ‘p’ seconds of each other in

C. Step Three

Count the number of such incidence and compare it with the total volume of flow on both stream flows. The value depicts the probability of a traffic crash. It is quantified as in (2)

$$P(\text{crash}) = \frac{\text{Total number of vehicles in conflict}}{\text{Sum of volumes of vehicle entering the facility}} \quad (2)$$

It is important to realize the type of conflict as this methodology only focuses on the probability and not the intensity of the crash. The type of conflict along with its probability can then be used to estimate the severity prediction of crashes and then ranking them.

VII. GIS AND THEMATIC REPRESENTATION

Conventional databases as well as the predicted probabilities lacked visibility, which is essential for better understanding and good decision-making. The fundamental objective of this paper is to determine a probability to accidents at those spots and to take actions that will reduce crash frequency or severity. Geographic Information System (GIS) has been identified as an excellent system for storing and managing these types of data and also as a potential tool for improving accident analysis process. One of the reasons is that it provides an efficient system of databases and provides a spatial referencing system for reporting output at different levels of aggregation.

The probabilities of crashes are thematically represented on the GIS road network of Karachi. Geomedia Professional, a

GIS software, was used to plot the probabilities of crashes. The facility under study is chosen, its satellite imagery procured, vectorized and finally cross checked for topological errors. Probabilities of crashes are imported directly from the spreadsheet files onto the GIS database. As the probability of crashes is a function of volume and time as well, thematic maps were generated for the 15 minute intervals to better visualize the scenario of that epoch. Alongside, a 16 hour accident probability theme was also generated to summarize and compare results. Outputs such as these were compared to current trend of accidents to check validity.

VIII. CONCLUDING REMARKS AND FUTURE RESEARCH

In this paper, a simple methodology was proposed in order to simulate and predict the probabilities of a traffic crash at a conflict zone. Importance was laid on the methodology to be low on cost. Therefore, stress had been laid on basic principles and assumptions and, thus, a probabilistic approach was adopted, requiring minimal data for analysis. Simulation was made using VBA scripts and spreadsheet applications which are a lot cheaper than professional traffic simulation software. These all insure small investment. Also, as no continuous data collection operation is required, operation costs are further reduced thus rendering the methodology to be financially optimal.

This methodology can adopt electronic data acquisition as well. Data collectors can save the traffic data file in a comma separated file (CSV) or similar format which could be imported on the spreadsheet for analysis. Furthermore, this methodology can accommodate continuous live data collection as well. By updating flow headways on the spreadsheet at very small time interval (e.g. 2 minutes) for the same 15 minute time interval, this methodology can accommodate the user with continuous probabilities thematically making the current methodology more adaptive and efficient.

Incorporation of driver behavior into the model would further improve upon the efficiency of this methodology. Alongside, GIS based analysis open doors to network based simulation. Incorporation of car following theory would further improve the flow prediction model. With these and the incorporation of ITS to such a model, Karachi would be stepping into its initial ATMS (Advanced Traffic Management System).

ACKNOWLEDGMENT

M. A. Rehmatullah thanks the help and support of Ashar Lodi, Director, Exponent Engineers for all the technical guidance and support necessary in completing this research endeavor. The author would also like to thank Prof. Sarosh H. Lodi for all the motivation and help without which this paper would not be possible.

REFERENCES

[1] Helman, D.L, in *Traffic Incident Management. Public Roads*, 68 (3), 2004, pp. 14-21.
[2] Bilham, R.,et. al., “Seismic Harzard in Karachi, Pakistan: Uncertain Past, Uncertain Future”, in *Seismological Research Letters*, vol. 78, No. 6, pp 601-613
[3] TRB, *Revised Traffic Flow Theory A State-of-the-Art Report*, 2001, ch. 2
[4] TRB, *Traffic Flow Theory, A Monograph*, 1975; ch. 3
[5] Tom V. Mathew et. al., Heterogeneous Traffic Flow Modeling and Simulation using Cell Automata, in *Applications of Advanced Technology in Transportation: Proceedings of the Ninth International Conference* , 2006, pp 492 – 197
[6] Nan Vou & Jianwei Wang, A Video Based Method for Evaluating Traffic Data from Detectors, in *Applications of Advanced Technology in Transportation: Proceedings of the Ninth International Conference*, 2006, pp 232 – 237.
[7] Adolf D. May, *Traffic Flow Fundamentals*, 1990, Prentice Hall, ISBN 0-13-926072-2
[8] Banks, J., et. al., in *Discrete Event System Evaluation*, 3rd edition Prentice Hall, 2001, pp. 269
[9] C.S. Papacostas, P.D. Prevedouros, *Transportation Engineering and Planning*, Prentice Hall Third Edition,
[10] McLeans, J.R., *Two Lane Highway traffic Operations – theory and practice*, Gordon and breach Science Publishers, New York, 1989.
[11] Weisstein, Eric W. "Monte Carlo Method." *From MathWorld - Wolfram Web Resource*. http://mathworld.wolfram.com/MonteCarloMethod.html