VCI Descriptive Model

Diogo Samuel Fernandes
Faculty of Engineering
University of Porto
Porto, Portugal
up201806250@up.pt

Juliane Marubayashi
Faculty of Engineering
University of Porto
Porto, Portugal
up201800175@up.pt

Paulo Ribeiro
Faculty of Engineering
University of Porto
Porto, Portugal
up201806505@up.pt

***Abstract*—As the social sciences and real-world structures increase in complexity, the modern community must find other alternatives to analyze and solve problems. Simulations arise in this context to find solutions and verify the consequences of possible actions. Given the high load of vehicles on Porto's main motorway (VCI) and its current problems, this paper describes how we developed a descriptive model of the VCI and analyzed how good our prototype is. Creating such a model eases future work on testing new hypotheses and validating solutions to improve the traffic on this motorway.**

I. Introduction

Porto is the city with the highest affluence of cars in Portugal. The increased traffic causes many problems, such as excessive pollution and stress.

In the context of smart cities, a large amount of data from the VCI (Via de Cintura Interna) motorway was gathered by inductive-loop sensors underneath the pavement all over the VCI. With this data, it is possible to recreate the observed and measured traffic on the streets in a simulation to test new ideas and solutions for improving traffic performance in real cities.

This paper is structured as follows. Section 2 presents the methodological approach followed to reach the results. Section 3 discusses the results. Section 4 indicates similar works to this one. Finally, Section 5 concludes with future work on the descriptive model.

II. Methodological Approach

A. Simulation Assumptions and Requirements

The simulation environment only considers cars from categories A and B, which can be generated at any entry point of this highway and end the simulation at any exit point. Another constraint in this model forbids cars from driving in the reverse direction.

B. Dataset Description

The data we obtained comes from inductive-loop sensors installed underneath the pavement all over the VCI inner ring in Porto, capable of gathering traffic information. Since this data was provided to us by the professor, no data collection steps will be necessary. However, we must process the data to use it as input for the model developed in SUMO. For this, it may be necessary to analyze all variables and their probability distributions, and a great understanding of their characteristics is essential for the project's success. The data provided corresponds to 2013 (although this year's data are incomplete due to concession problems), and we also have access to both the 2014 and 2015 semesters. These are aggregated data in five-minute periods. In other words, for each interval of that time, we have access to an aggregated calculation that gives us, for example, the total number of cars that passed by the sensor in that period or the average speed. The data contains many attributes, including the number of vehicles of each class (A, B, C or D) that passed through the sensor, from which we only consider the first two.

Here we highlight the most relevant ones for our project:

- *EQUIPMENT_ID*: the sensor's unique ID;
- *LANE_NR*: the current lane number where the data is measured;
- *LANE_DIRECTION*: the current lane direction where the data is measured. May take the values "C" or "D", which mean ascending and decreasing, respectively;
- *VOLUME_CLASS_A*: Number of vehicles of class A that pass through the sensor;
- *VOLUME_CLASS_B*: Number of vehicles of class B that pass through the sensor;
- *AVG_SPEED_ARITHMETIC*: Arithmetic average speed of vehicles that pass through the sensor;
- *AVG_SPEED_HARMONIC*: Harmonic average speed of vehicles that pass through the sensor;
- *AGG_PERIOD_LEN_MINS*: time between each measurement the sensors made (in this case, it is always five minutes);
- *AGG_PERIOD_START*: time interval when the data was collected, with the format "Year-Month-Day Hour:Minutes:Seconds";

Additionally, we also have access to the properties of each sensor, like the identification of the street it covers. Nevertheless, for this project, it was enough for us to use the following attributes:

- *EQUIPMENT_ID*: the sensor's unique ID, allowing us to establish relationships with the traffic data files;
- *LATITUDE*: the latitude of the sensor;
- *LONGITUDE*: the longitude of the sensor;

This information allows us to locate the sensors on the map to create a digital version of them, which is necessary to collect the simulation data.

C. Modelling Approach

We divided the modelling approach into three main processes:

- Generate the routes for each origin-destination pair;
- Generate the origin-destination matrix containing all the possible entering and exiting nodes of the map;
- Calibrate the current model.

Each entry in the OD matrix is represented as a triple $(O \times D \times N)$. N is the number of cars with O as the origin and D as the destination node. To avoid repeating the task numerous times, routes for each (o, d) pair were pre-calculated with the sumolib method (i.e., getShortestPath) and saved in a file.

To generate the OD matrix M, the model gathers the real data and stipulates the number of cars that go from a node o to a destination d. Firstly, we processed the real data by calculating the average volume of vehicles per hour for each sensor between 8 a.m. and 10:30 a.m. in 2015. The average volume of each sensor i located at an edge e_i was equally distributed among each pair $(o_j, d_j) \in M$, where $e_i \in E_j$. Therefore, the value N for an arbitrary pair (o, d) is equals to:

$$\sum_{0}^{i=n} \frac{v_i}{m_i}$$

Where n is the number of sensors that pass through the route of (o, d), v_i is the average volume of the sensor i, and m_i is the number of paths that the sensor passes through.

The main goal of the calibration step is to run the simulation, compare it with the produced values in the real data and calculate the error as approached in the next section for the many different scenarios described at III.

D. KPIs

For the calibration process of our model, it will be necessary to define a Key Performance Indicator, a measurable value that demonstrates how effectively the model represents the actual traffic on VCI. This metric could consist of an error that reflects the differences between the real data and the data produced by the simulation. Our main objective would be to reduce this value as much as possible, bringing our model closer to reality and thus achieving the goal of creating a reliable descriptive model of the VCI. It is necessary to decide which traffic attributes to consider. We base the error calculation on three of them: the total volume of cars that entered and left the sensor road during the observation, the arithmetic average speed, and the harmonic average speed. We used the mean absolute error, which corresponds to the sum of the absolute values of the differences between the data and is given by:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}$$

In this formula, n is the number of observations, y is the real sensor values, x is the SUMO sensor values, and e is the resulting difference between both values. We chose this error measure because it is less sensitive to outliers, which are very frequent in inductive loop data detectors, due to the difficulty of obtaining accurate and noise-free data from the sensors.

III. Results and Discussion

A. Scenarios

To calibrate the model, we followed an optimization process through simulation, where we tested various scenarios to find out which was closest to the actual traffic behavior on the VCI.

As explained in subsection II-C, before each simulation, we define an enumeration of the existing OD pairs and, for each, the number of cars that should depart from that origin and leave at that destination. After creating the default matrix M described in that subsection, we produced five scenarios, which proportionally reduced the number of vehicles in the simulation.

Although SUMO is a powerful tool, we assumed that the high number of cars in the simulation could cause overhead and increase the error. Considering that M contains 100% of the vehicles that were supposed to be on the simulation, we have created other matrices containing 10%, 25%, 50%, and 75% of the cars: M_{10}, M_{25}, M_{50}, and M_{75}.

We then run a simulation for each of these matrices and store the error value returned. In subsection III-C, we describe these results in detail for both assessment levels previously described (five-minute blocks and one-hour blocks).

B. Experiments

To determine the accuracy of the simulation results, two main methods for calculating errors are used. The first approach involves comparing the output from the simulation software (SUMO) with each individual record of the real data (five-minute blocks). The second method involves calculating the volume of cars in each simulated sensor per hour (grouping data into one-hour blocks), and then comparing that result with the same calculation performed on the real data. This second approach provides a less demanding comparison because the accumulated error is more significant in blocks of five minutes. The next subsection describes the results of both these approaches.

C. Results

We performed a thorough evaluation of various traffic scenarios on the VCI to determine which scenario closely mimics real-world traffic behaviour. The study was designed to measure the realism of the traffic scenarios by analyzing the maximum value in each scenario, which represents the total number of vehicles resulting from the equal distribution of flows from each sensor among the Origin-Destination (OD) pairs whose routes pass through them. The results of this analysis can be seen on Table I and Table II.

IV. Related Work

In the literature, other investigations are related to VCI (e.g., [1] and [2]). In contrast with our project, the previous studies mainly focus on how to collect data from roads and

TABLE I
RESULTS OF FIRST STRATEGY

10%	25%	50%	75%	100%
47,576.070	46,954.375	46,835.290	47,223.463	47,029.250

TABLE II
RESULTS OF THE SECOND STRATEGY

10%	25%	50%	75%	100%
4,671.672	4,485.051	4,412.266	4,346.030	4,193.740

how to extract knowledge. Although different strategies were applied, this paper shares the same goal from previous works: improving the VCI traffic system.

V. CONCLUSION

This project creates a descriptive model of the VCI, to test new scenarios, understand its current problems and prescribe new solutions for future improvement. We've successfully applied modelling methodologies to enhance and validate the descriptive model. Although the error is still significant, by exploring more scenarios, the variation of the error will follow a diminishing marginal return and stabilize at some moment. Future work will involve more in-depth testing of different scenarios in order to bring the model even closer to reality. To complement this work, prescriptive and predictive models could also be built, which could be useful in traffic forecasting, providing important information for traffic control centres.

ACKNOWLEDGMENT

This report contains the work done in the curricular unit Modelling and Simulation lectured by professor Rosaldo Rossetti at the Faculty of Engineering of the University of Porto.

REFERENCES

[1] I. Alam, D. Md. Farid, and R. J. F. Rossetti, “The Prediction of Traffic Flow with Regression Analysis,” in Emerging Technologies in Data Mining and Information Security, 2019, pp. 661–671.

[2] Alam et al., ”Pattern mining from historical traffic big data,” 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, India, 2017, pp. 1-5, doi: 10.1109/TENCONSpring.2017.8070031.