ON DATA ACQUISITION, MODELING AND SIMULATION OF HIGHWAY TRAFFIC

Tom Bellemans * **Bart De Schutter** ** **Bart De Moor** ***

* *ESAT/SISTA, K.U.Leuven, Kard. Mercierlaan 94, B-3001 Leuven (Heverlee), Belgium, tom.bellemans@esat.kuleuven.ac.be*
** *Control Lab, Fac. ITS, TU Delft, P.O.Box 5031, 2600 GA Delft, The Netherlands, b.deschutter@its.tudelft.nl*
*** *ESAT/SISTA, K.U.Leuven, Kard. Mercierlaan 94, B-3001 Leuven (Heverlee), Belgium, bart.demoor@esat.kuleuven.ac.be*

Abstract: We consider the process of setting up a traffic simulation based on traffic measurements. This process consists of several steps: data acquisition, interfacing, data conditioning, modeling and simulation. We discuss some important aspects that should be taken into account in each step of the process, and we also illustrate the entire process using a real-life traffic example.

Keywords: data processing, data acquisition, preprocessing, computer simulation, filtering, interpolation, road traffic

1. INTRODUCTION

Highways nowadays play an increasingly important role in our transportation system. However, especially during rush hours, many drivers are confronted with traffic jams, leading to lost work and leisure time, increased fuel consumption, air pollution, health problems, stress, and discomfort. Therefore, the prediction of traffic jams and the determination of the effectiveness of various traffic control measures to reduce traffic jams and to obtain smoother traffic flows are an important topic. One of the tools that can be used to answer these questions is traffic simulation.

In this paper we illustrate the different steps of the process in which a simulation model is constructed based on on-line measurements of traffic parameters. The different steps of the process are schematically presented in Fig. 1. We discuss some important aspects that should be taken into account in each step of the process, and we illustrate them by means of a particular case study in which we have modeled the traffic flows on the ring-road around Antwerp, a major Belgian city. We also discuss the transformation of the raw sensor data to the standard format of the central database using data filters. Furthermore, we present a method to obtain good estimates of missing data values in a computationally efficient way. Next, we discuss an efficient on-line method to estimate the input parameters for the simulation. Finally, we consider a microsimulation of a real-life traffic situation in Antwerp (Belgium) based on the process described in this paper.

2. DATA ACQUISITION

2.1 *Measured parameters*

When we want to observe or model a traffic situation based on measurements, some important parameters are (May, 1990): number of vehicles per time unit (traffic intensity), speed of the vehicles, occupancy level of the highway, time headways, distance headways, and traffic density.

The most important parameter is the number of vehicles passing at a certain point in a certain

Fig. 1. Schematic overview of the steps necessary to simulate a real-life traffic situation.

time interval. This is a good measure for the traffic intensity on the highway. The (average) speed of the vehicles is another important parameter to assess the traffic situation on the highway. The occupancy level is the fraction of the total time the detector detects a vehicle. The time headway is the time that passes between the arrival of the front of a first vehicle and the front of a second vehicle at a reference point. The distance headway is the distance between the front of the first vehicle and the front of the second vehicle.

Depending on the set-up of the measurement system, we can also distinguish several levels of detail: We can consider average values (time-averages for a given location, or averages for a given section of the highway) or we can record instantaneous values for each vehicle. The traffic parameters can be measured for each lane separately or for the highway as a whole. Furthermore, a distinction can be made between different classes of vehicles, e.g., cars, lorries, HOVs. As a rule of thumb, we can say that the higher the level of detail required and the more parameters have to be measured, the higher the cost of the measurement system will be.

2.2 *Sensor systems*

There are several technologies to measure the traffic parameters described in the previous section (Johnson, 1999): pneumatic detection systems, inductive loops, cameras, ultrasonic sensors, microwave sensors, active and passive infrared sensors, passive acoustic arrays, and magnetometers. In our project we have used pneumatic detection systems, inductive loops, and cameras. Now we briefly discuss the advantages and restrictions of these traffic sensor technologies.

In a pneumatic detection system, a tube is led over the road. A car running over the tube produces a shock wave in the tube, which can be detected. The main advantage of this technology is the mobility of the installation. It allows for a quick set up of a measuring point along a highway. A disadvantage is the limited accuracy (e.g., it is difficult to tell whether two successive cars passed over the cable or a lorry with four axles). Another disadvantage is the fact that no distinction can be made between cars on different lanes. Pneumatic detection systems are in general only used to count the number of cars that pass the detection point; speeds and occupancy levels are difficult to determine using the pneumatic sensors.

Inductive loops consist of a conducting loop that is put in or on the surface of the highway. This loop is connected to an electronic circuit which produces an oscillating current through the loop. If a car enters the loop, the metal parts of the car cause the oscillation frequency of the electronic circuit to shift. This frequency shift can be detected. There are two main configurations of inductive loop systems: the single-loop and the double-loop configuration. Using the single-loop configuration, we can easily obtain the number of vehicles passing and the occupancy level of the highway for each lane separately, which is the main advantage of this technology. However, using a single inductive loop set-up and an average vehicle length, we are not able to distinguish a fast driving lorry from a slow driving car, since the time the vehicle is in sight of the detector (in the loop) can be the same. In order to solve this problem and to improve the accuracy of the measurements, we can use a double configuration of inductive loops. The main advantages of the double inductive loop set-up are the improved accuracy and the diversity of traffic parameters that can be measured. The disadvantage is the higher cost with respect to the single loop configuration. The main disadvantages of loop techniques are that the loops become easily damaged during road works and that if the loops are put in the road surface then installation and maintenance require special measures and thus also extra costs.

A more recent technology is the use of cameras and image processing algorithms. The set-up is as follows. A camera is placed above the highway. The image of this camera is sent to an image processing unit which extracts the traffic parameters using image recognition algorithms. The image of the cameras can also be used as a visual inspection tool as can be seen in Fig. 2. The main advantages of this set-up are the higher accuracy of the measurements and the visual inspection possibilities, but the main disadvantage is the higher cost.

In general we can say that inductive loops provide better accuracy than pneumatic detection systems and that they are cheap compared to the camera technology and the other advanced sensor technologies mentioned above (although the latter may offer a higher accuracy).

3. INTERFACING

In our set-up the sensors are connected to the central computer via a network. The central computer regularly polls the sensors for their measurements. The typical time interval between two polls is one minute, but for some applications such a frequent update is not necessary and another schedule can be used (e.g., poll every 15 or 30

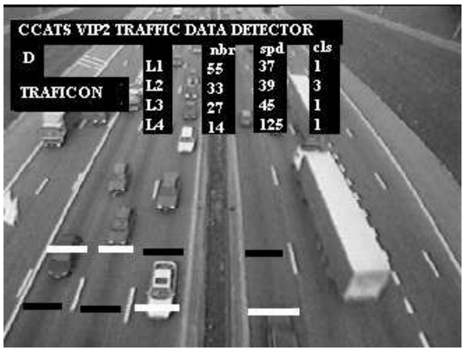

Fig. 2. Camera view of a highway, used to measure the traffic parameters of the different highway lanes. (Photo courtesy of Traficon)

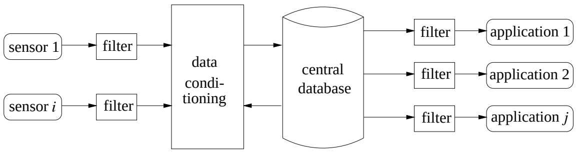

Fig. 3. Schematic view of data flows from sensors towards the database and from the database towards applications.

minutes). After the data are received, the central computer stores the data in a database in the memory or on disk.

In order to make an efficient access or download of the data possible, we use a central database. For this central database we define a standard data format which must allow for an easy exchange of data. In general the data generated by different sensors have different formats. Therefore, the measurements from the sensors are converted into the standard format using data filters. A big advantage of this approach is that it improves the modularity of the set-up as is illustrated in Fig. 3. If a sensor is replaced by another brand or technology, we only have to change the data filter. The database and its components remain unchanged. This is also the case for the applications using the data of the database: a changed or new application can be made operational by simply applying a new filter. This requires no changes to the central database system nor to the application itself.

4. DATA CONDITIONING

4.1 *Corrupt and missing data*

When the sensor data is transferred over a network to a central computer, there is in general a check on the values of the data. For every measured parameter constraints can be defined based on physical laws, and common-sense assumptions, e.g., a vehicle count cannot be negative, the speed of a car on a highway cannot be negative and cannot be larger than 200 km/h. If a measured value does not satisfy the constraints, it is marked as unavailable by writing a sentinel element in the database. In that way the corrupt data values are transformed into missing data values.

Besides the sentinel elements introduced by the check on the physical constraints, there are also other causes of missing data. Sometimes the central computer gets no answer to its polls because a sensor is out of order due to works on the road, power failure, ... or because the network is not working properly. Since most simulation packages require data records which are complete and contain no gaps, we discuss some methods to estimate the missing values.

4.2 *Estimation of missing data*

When a measurement value is missing, it can be estimated using different estimation techniques. In general we start from the values that are available and we construct a function that approximates the given data points. For the approximation function we can use piecewise-linear functions, polynomial functions, splines, and so on. Next the missing values can be determined by interpolation or extrapolation.

The main advantage of linear interpolation or extrapolation is its computational efficiency. The main disadvantage is its limited accuracy especially if we use extrapolation, i.e., if the time index of the missing values lies outside the time range of the available data points. This becomes important when a measurement is unavailable for a longer time, since then estimates are needed for successive values. Interpolation and extrapolation are in fact local estimation methods which use a linearization in the neighborhood of the available measurements. Usage of this linearization over a wider interval can give rise to large deviations of the estimation from the real value. Therefore, we can try to use a better approximator such as polynomial approximation or spline-based approximation. The use of more data points results in a higher attainable accuracy, but also increases the computational efforts. Next, we present an estimation scheme which takes this kind of trade-off into account.

4.3 *Problem specific estimation scheme*

The method we have used in our project boils down to an interpolation of the missing value with a reference day, which is constructed off-line using many data points. The reference day is a day whose traffic situation is representative for the day under study. The missing values at a given time instant of the current day are determined based upon the last available measurement and information about the traffic situation at the corresponding time instant of the reference day. The missing value is looked up in the reference day and scaled to the traffic intensity of the studied day. This scaling is done using the following formula:

$$x_{\text{data}}(k) = \frac{x_{\text{data}}(k-1)}{x_{\text{reference}}(k-1)} x_{\text{reference}}(k)$$

where $x_{\text{data}}(k)$ is the measurement (or the estimate) of the variable at the kth sampling time instant of the current day, and $x_{\text{reference}}(k)$ is the corresponding value in the reference day.

The idea behind this method is that every day will show a traffic pattern which can be predicted, e.g., the build-up of congestion in the morning and the evening rush hours. If we can find a day with the same traffic pattern, which we call the reference day, we can use its information about the traffic pattern to estimate the missing values. This is done by scaling the value from the reference day to the intensity level of the current day using the scaling factor $\frac{x_{\text{data}}(k-1)}{x_{\text{reference}}(k-1)}$. This scaling is necessary since the traffic intensities of the present day and the reference day are not necessarily the same in absolute numbers whereas their traffic patterns (i.e., the evolution of the traffic intensities) are assumed to be identical. The reference day is constructed by taking the average over several days which we expect to have the same traffic pattern. For each time instant the average is taken by first adding the corresponding available measurements and then dividing by the number of available values in the sum. This way, it becomes easier to construct a reference day, since we can also use days with some missing measurements to construct a reference day. This method also averages out incidental fluctuations, which leads to a more dependable reference day and to more robust estimates.

The traffic patterns during a holiday are completely different from the patterns of an ordinary working day, even if both days are the same week days. Also, certain happenings (pop concert, football game, ...) or incidents (accident, works on the road, ...) can influence the traffic intensity. Last but not least, the weather also plays an important role in the way a traffic situation develops. Therefore we must avoid to include days of which we can expect a special traffic pattern into the calculation of the reference day. It is clear that a method which averages over several days will be more robust against deviations.

If we use a long averaging window, which means averaging over many days, deviating values will be suppressed. We could, e.g., take the average over all Mondays for a year to get a reference day to use for Mondays. This is not a realistic situation since we then have the same reference day for a Monday in summer and a Monday in winter, even though the seasons influence the traffic patterns. We see that there is a trade-off between the averaging out of stochastic effects and tracking behavior. If we make the averaging window smaller, e.g., two months, which corresponds to eight Mondays, we will be able to track changes caused by the seasons.

We have used the method with the reference day in our project since it performs better in cases where several values are missing, which is in general the case in our application. Moreover, once the reference day is used, this method is computationally as efficient as ordinary interpolation or extrapolation while giving more accurate results. So the reference-day method offers a good trade-off between accuracy and computational efficiency.

5. MODELING OF HIGHWAY TRAFFIC

We could use analytic methods or computer simulation methods to describe and analyze the traffic flows on the highway (de Dios Ortúzar and Willumsen, 1994; Gartner *et al.*, 1997; Newell, 1980). We have used a computer simulation model in our project. We can distinguish two large classes of simulation models, based on their level of detail:

- Macroscopic models or macro models describe the evolution of the traffic in terms of flows and densities (Chen *et al.*, 1997). In general a macro model uses a discretization in time and space of the highway network under consideration. The model is built with different sections (discretization in space) which are coupled together to form the studied layout. The state of each section is updated every iteration (discretization in time) using the state equations that relate the next state of the system to the current state and inputs of the section. Typical state variables of the section are the vehicle density, the average velocity and the traffic flow through the section. The coupling of the sections is realized by considering the out-flow of a section as the in-flow of the next section.
- Microscopic models or micro models (see Fig. 4) simulate the highway network on the level of the individual vehicles. The physical layout of the highway network is entered into the simulation engine. The vehicles move around in the network taking into account their properties and the restrictions caused by the section and the other vehicles. Every vehicle can be followed on its way through the network.

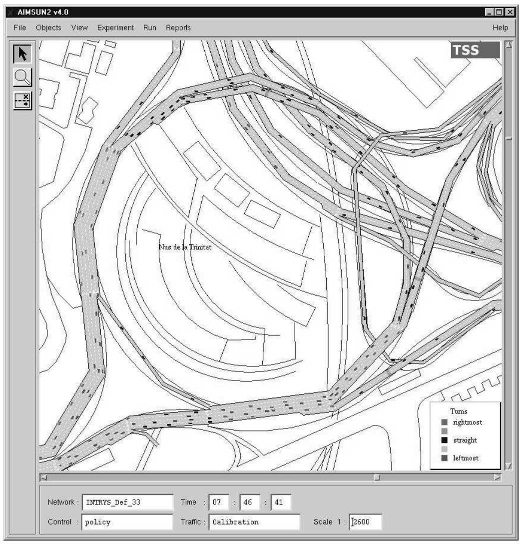

Fig. 4. Screen-shot of a micro simulation. (Photo courtesy of TSS)

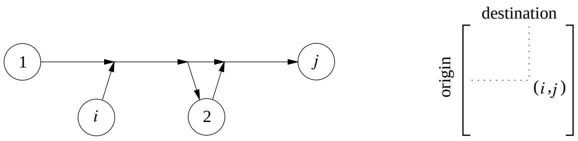

Fig. 5. An OD matrix describes the flows in the network. The ijth entry of the OD matrix is equal to the number of vehicles that go from zone i to zone j.

Micro models are very detailed, but they are computationally very intensive and show stochastic behavior, whereas macro models are less detailed but require less calculations, which makes them more suitable for, e.g., real time applications. In general we could say that the choice of model is dependent on the problem we are faced with. If a fast model is needed, we choose a macro model, but if much detail is required and enough computational power is available, we can use a micro model. In our project we have used a micro model since we wanted to gain a detailed insight in the traffic flows on the highway network under consideration and since sufficient computational power was available.

6. SIMULATION OF HIGHWAY TRAFFIC

Simulation is the last step in our process. An important input parameter for the traffic simulator is the origin-destination matrix (OD matrix) of the system, which describes the flows between the different zones in the network.

6.1 *Estimation of OD matrices*

Based on the traffic measurements we try to find an estimation of the traffic flows in the network. These flows are collected in an origin-destination matrix or OD matrix. The ij-th element of the OD matrix corresponds to the number of vehicles traveling from zone i to zone j of the network (see Fig. 5). This construction results in an OD matrix the entries of which are the flows in the network in absolute numbers. Since the OD matrix is needed as an input for the simulator, we now discuss the method we have used to make an estimation of the flows between the different zones based on the measurements available in the database. Note that the obtained solution will only yield an estimate of the real flows between the zones since in general measurement errors occur and since multiple flow patterns may exist that correspond to the same measurements. The estimation of the OD matrix is made based on two types of equations: hard equations and logical equations:

- Hard equations are imposed by the layout of the network, e.g., if two zones are not connected, the flow between these zones is zero. Another example concerns the diagonal entries of the OD matrix. These are zero since the number of vehicles traveling through the network from zone i to zone i is zero.
- The hard equations alone are not sufficient to uniquely determine the flow pattern. Therefore we have added some logical equations in order to reduce the set of possible solutions and to find a solution which will better correspond with reality. These equations are not imposed by the layout of the system but by the behavior of drivers. A typical example of a logical equation is that it is highly improbable that a driver that enters a highway will leave it on the next off-ramp (since in our case study off-ramps and on-ramps are located at short distances from each other).

Both the hard equations and the logical equations are in general linear constraints of the form $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{x}$ contains the entries of the OD matrix. Besides these equations, there is also an important constraint added by the nature of the problem: since we are dealing with flows, the values must be nonnegative, i.e., $\mathbf{x} \geq \mathbf{0}$. However, we also have to deal with measurement errors which cause the hard equations and the logical equations not to hold exactly (i.e., in general we have $\mathbf{Ax} \approx \mathbf{b}$ instead of $\mathbf{Ax} = \mathbf{b}$). Therefore, we try to find the best solution according to a least-squares criterion. This leads to the following nonnegative least squares problem:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2 \qquad \text{subject to } \mathbf{x} \geq \mathbf{0} \ ,$$

where $\|\mathbf{y}\|_2^2 = \mathbf{y}^T\mathbf{y}$ for any vector $\mathbf{y}$. This problem can be solved using the nonnegative least squares algorithm of (Lawson and Hanson, 1995).

Other and more extensive methods to estimate OD matrices are described in (van der Zijpp and Hamerslag, 1994). However, since in our application the estimate of the OD matrix has to be updated regularly in order to take the variation in time of the traffic patterns into account, we have selected the method described above since it offers a good trade-off between accuracy and computational efficiency.

6.2 *Traffic simulation*

When the OD matrices are available, we can use them to do a simulation of the traffic situation. The simulations are done with a simulator based on dynamic traffic models such as PARAMICS, AIMSUN2, MITSIM, etc.

We now illustrate the process described above with the simulation of a real-life traffic situation. We have simulated the traffic on a 7 km stretch of the E17 highway near Antwerp (Belgium) based on the measurements available from the existing measurement system. On the highway stretch cameras are available every 500 m over a length of 7 km. These cameras provide us with the number of vehicles per minute and per lane. We also have the average velocity and the occupancy rate available for every lane. The values measured by the cameras are sent over the network after being processed by a standardization filter. In the data conditioning step the measured densities are compared to upper and lower thresholds in order to remove corrupt values. Next the missing values are estimated using the reference-day estimation method of Section 4.3. Based on these data, we start our OD estimation routine. The OD matrix is re-estimated every 15 minutes. We now can run a simulation of the network. This simulation can be used to make a prediction of the future evolution of the traffic situation, and to evaluate the impact of various traffic control measures (such as ramp metering, addition or removal of lanes, changes in the layout of the highway) on the traffic flows.

7. CONCLUSIONS

We have given an overview of the steps needed to set up a traffic simulation where we have used a specific case study to illustrate the process. On the lowest level we find the sensors which measure certain traffic parameters. These measurements are transported to a central computer and stored in a central database. In order to obtain a modular structure in which each component can easily be replaced, we have used an architecture based on data filters and a standardized data format in the central database. We have also discussed how corrupt and missing values can be dealt with. Next we have presented a method to obtain estimates of missing data. This method is based on a reference day, and offers a good trade-off between accuracy, robustness and computational efficiency, which is very important for the kind of on-line applications used in our case study. We have used a microscopic computer simulation model to model the traffic flows. An important input parameter for the simulation is the origin-destination matrix (OD matrix). We have also presented a method that allows us to estimate OD matrices from the traffic measurements on-line. Finally, we have briefly discussed a simulation of a real-life traffic situation.

Acknowledgments

Bart De Moor is a senior research associate with the F.W.O. (Fund for Scientific Research-Flanders).

This work is supported by several projects: GOA Mefisto-666, FWO projects G.0240.99 & G.0256.97, FWO-Research Communities ICCoS & ANMMM, IWT projects EUREKA 1562-SINOPSYS, EUREKA 2063-IMPACT, & STWW, IUAP P4-02 and IUAP P4-24, Sustainable Mobility Programme-Project MD/01/24 (1997-2000) "Traffic Congestion Problems in Belgium: Mathematical Models, Simulation, Control and Actions", TMR Networks ALAPEDES & System Identification.

8. REFERENCES

Chen, O.J., A.F. Hotz and M.E. Ben-Akiva (1997). Development and evaluation of a dynamic ramp metering control model. In: *Proc. of the 8th IFAC/IFIP/IFORS Symposium on Transportation Systems*. Chania, Greece. pp. 1162–1168.

de Dios Ortúzar, J. and L.G. Willumsen (1994). *Modelling Transport*. 2nd ed. Wiley.

Gartner, N.H., C.J. Messer and A.K. Rathi (Eds.) (1997). *Monograph on Traffic Flow Theory*. Oak Ridge National Laboratory (ORNL). See also http://www-cta.ornl.gov/cta/research/trb/tft.html

Johnson, B. (1999). Keeping the world flowing. *Traffic Technology International* pp. 38–40.

Lawson, C. and R. Hanson (1995). *Solving Least Squares Problems*. Vol. 15 of *Classics in Applied Mathematics*. SIAM. Philadelphia, Pennsylvania.

May, A.D. (1990). *Traffic Flow Fundamentals*. Prentice-Hall. Englewood Cliffs, New Jersey.

Newell, G.F. (1980). *Traffic Flow on Transportation Networks*. MIT Press. Cambridge, Massachusetts.

van der Zijpp, N.J. and R. Hamerslag (1994). An improved Kalman filtering approach to estimate origin-destination matrices for freeway corridors. *Transportation Research Record* (1443), 54–64.