Identification of parameters and verification of an urban traffic flow model. A case study in Wrocław

Mieczysław Chalfen¹*, Joanna A. Kamińska¹

¹Wrocław University of Environmental and Life Sciences, Department of Mathematics, ul.Grunwaldzka 53, 50-357 Wrocław, Poland

Abstract. A macroscopic model of road traffic flow in an entire city is constructed, using the example of Wrocław. The model is of deterministic-random type. The start and finish points of each vehicle journey are random elements of the model, while the street graph, routes of travel and traffic parameters are obtained in a deterministic manner. Vehicle speed is dependent on traffic density, and the time needed to cross an intersection depends on the number of waiting vehicles. The route of travel between given start and finish points is determined using Dijkstra’s algorithm, minimising the journey time. The street graph of Wrocław was constructed using data from the Open Street Map website. Identification of parameters and verification of the model were performed using hourly data on traffic volumes at the city’s major intersections in the years 2015–16, obtained from the Intelligent Transport System (ITS). Identification was made of the total number of vehicles travelling in the city at each time of day, a quantity that is difficult to determine by other methods. An accuracy of around 15% was obtained for times between 6.00 and 22.00, while for night-time the verification error exceeded 30%. The model may be used to analyse the impact of planned modifications to the transport system on traffic parameters in the city. It may also serve as a constituent of a larger model for investigating the effect of road transport on atmospheric pollution or for identifying areas at risk of noise pollution.

1 Introduction

The modelling of vehicular traffic flow is a method commonly used to identify transport-related phenomena occurring in urban road networks. Three basic classes of road traffic model may be distinguished: macroscopic, mesoscopic and microscopic. Microscopic models serve to study the behaviour of up to 10 or 20 vehicles, and are applied to small-area transport systems limited to one or several neighbouring intersections. Mesoscopic models reflect a medium level of detail: the movement of individual vehicles is analysed, but their behaviours on traffic lanes are generalised by aggregating vehicles into groups and by analysing only the time and not the exact course of vehicle movement. Macroscopic models describe the movement of a stream of vehicles in a complex road network covering a whole city, where the movement of each vehicle is reflected separately but in a simplified manner. Several different methods are used to model the movement of vehicles [1], including methods based on the differential equations of gas kinematics [2], fluid dynamics [3-7], analogy with groundwater flow [8], the use of cellular automata [9-10], finite element methods [11, 12] and the application of graph theory. The model constructed in the present study is a macroscopic model based on graph theory and probability theory. Graph theory, with the application of Dijkstra’s algorithm, is employed to determine the route represented by the shortest (in a defined sense) path in a graph. The route choice problem has been widely analysed in the literature [13, 14]. The complexity of the process of determining the optimum path results from the need to determine simultaneously the journey costs, drivers’ preferences, and available information on traffic volumes, together with estimation of the use of that information by drivers. Journey costs may be defined in terms of the time taken to cover a given route, the length of the route, or a combined value (time costs + fuel costs + others). The first two possibilities lead to easy calculations, since they are deterministic in nature [15], but this is not true of the combined approach, since drivers have different preferences: time savings, short routes, easy routes, favourite routes [16]. In this study routes were determined so as to minimise the journey time. Probability theory was used to define probability distributions for the random sampling of journey start and end points. In the process of identification and verification of model parameters, account was taken of the effect exerted by current information about city traffic volumes on the choice of route at the time when a vehicle begins its journey. The question of the effect of information on the selection of a route, although related to route choice, constitutes a separate problem [17-19]. Lindsey [20] studied the effects of pre-trip information on route-choice decisions when travel conditions are congested and stochastic. In the study [21] it was shown that access to and use of continuous information on

*Corresponding author: mieczyslaw.chalfen@upwr.edu.pl

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).
traffic density may, by way of mid-journey route modifications, reduce travel times by 4% in very congested traffic and by as much as 24% in light traffic. There are many commercially available simulation packages which take account of vehicle movements on scales ranging from micro to macro [22, 23]. The best-known example may be the VISSIM system for the simulation of road traffic flow [24].

2 Model description

The traffic flow model for Wrocław presented here is a macroscopic model, dynamic in time and continuous in space. It consists of three basic elements: a graph of the city’s streets, a method of random sampling of journey start and end points, and an algorithm for determining the route taken by each vehicle. Any value may be taken for the time step size (the step size used in the simulations was Δt = 1 s). At each time point the exact position of all vehicles moving along the edges of the graph is determined. The model includes many traffic parameters which are hard to determine precisely. Some of them are estimated on the basis of literature data (distance between moving vehicles, mean vehicle length, duration of traffic light cycles at intersections), some based on the authors’ observations (very common small violations of the speed limit), and some by comparing simulation results with observations of actual traffic volumes (number of vehicles making journeys in the city as a whole). The model constructed is a macroscopic model, and hence it is less important to reflect faithfully the movement of a single vehicle. The aim of the model simulations is to obtain a correct description of traffic volumes and densities – which are variable in time and space – on the scale of the city as a whole, depending on various conditions and random events.

2.1 The street graph

The area covered by the simulations (Fig. 1) is bounded to the west by the starting point of the currently constructed Leśnica bypass, to the north by the Długoleka link, to the east by the Wrocław Eastern Bypass, and to the south by the currently constructed Southern Bypass up to the Bielany junction and part of the A4 motorway. This area contains the whole of the city of Wrocław together with the western bypass (S5) and its junction with the motorway. A graph of Wrocław’s streets was constructed using data available on the website www.openstreetmap.org, by selecting from the public database the nodes and edges tagged as ‘primary’, ‘primary_link’, ‘secondary’, ‘secondary_link’, ‘trunk’, ‘trunk_link’, ‘motorway’, ‘motorway_link’, ‘tertiary’ and ‘tertiary_link’. Local neighbourhood streets, of low significance for modelling traffic flow on the scale of the city as a whole, were ignored (Fig. 1a). The degree of detail used in representing the road network is illustrated in Fig. 1b. To reduce the large number of nodes, those nodes separated by a distance of less than 10 m and not representing road bends or corners were ignored.

Fig. 1a. The street graph (light-controlled intersections are marked in red).

Fig. 1b. Fragment of the graph representing the AOW–Kosmonautów intersection.

For every edge linking two vertices, the attributes ‘maxspeed’, ‘lanes’, ‘oneway’ and ‘name’ were read from the database. These indicate the maximum speed limit, the number of traffic lanes in each direction, whether the street is one-way, and the name of the street. The street graph finally obtained includes more than 8000 nodes, of which 570 are controlled by traffic lights. The network of streets has a total length of 765 km.

2.2 Start and finish points

Journey start points (S) and finish points (F) are a random element in the model. The spatial distribution of these points across the city changes depending on the time of day. It was decided that between 5.00 and 14.00 start points would be sampled according to a map of population density (Fig. 2), and finish points according to a two-dimensional normal distribution with a maximum in the centre of the city. This corresponds to a situation where, in the morning and until midday, drivers set out from home, and their destinations are
spread randomly over the area of the city, with a concentration of points in the centre. At other times the assumptions were reversed: start points were randomly distributed over the city, and finish points were sampled according to the population density map [25].

In the case of Wrocław, where the regions of highest population density are located chiefly in the centre of the city, such detailed differentiation of the location of start and finish points does not have a large effect on the results obtained in the traffic density simulations. The number of journeying vehicles, which is variable in time and not known a priori for each time of day, was determined by parameter identification in a subsequent part of the work.

Fig. 2. Population density map.

### 2.3 Route choice strategy

In practice, drivers setting off from a journey start point select their route according to their own knowledge and experience. Some drivers choose the shortest route, some choose the fastest, and some select only main roads, avoiding intersections where they would have to give way. The strategy selected by a driver is another random element in the model which is difficult to reflect precisely. In constructing a model it is possible to assume that each driver, when beginning a journey, selects the route that is best for him or her – meaning either the shortest in terms of distance, the fastest in terms of time, or the cheapest according to a defined cost function. In the remainder of the work it will be assumed that each driver, when setting out, chooses the fastest route in terms of time based on the traffic conditions existing at the time of the start of the journey. Models may allow for the possibility of a driver’s changing the previously planned route mid-journey for another more favourable route [21], but in the present study it is assumed that such behaviours do not occur.

### 2.4 Journey time

The time of a journey from the start point to the finish point along a pre-planned route is the sum of the times taken to travel along successive street sections, represented by edges of the graph, and the times taken to cross intersections, represented in the model by nodes. These times depend on the current road conditions, traffic volume and density, the number of vehicles waiting at an intersection, limitations on the vehicle’s speed, and the speed limit on particular road sections.

#### 2.4.1 Time of travel between nodes

The speed of travel along each street section is dependent on the density of traffic on that section at a given time, namely the number of vehicles per kilometre divided by the number of traffic lanes on that section. The speed is computed from the formula (1):

$$v(k) \leq \frac{1}{k} \left( \frac{3600}{t_p} - \frac{3.6d}{t_s} \right)$$

where:
- $k$ is the current traffic density [veh/km/line],
- $t_p$ is the safe distance [s],
- $d$ is the mean vehicle length [m],
- (it is assumed in the calculations that $t_s=2s$, $d=5m$).

The formula (1) was derived theoretically and verified in [26]. The computed theoretical speed is constrained by the speed limit applicable on the road section and by the vehicle’s maximum speed. In reality, however, not all drivers travel at the maximum speed determined from inequality (1). Some drive more slowly, while some slightly exceed the permitted speed limit. A model representing the actual movement of vehicles in the city ought to take account of such cases. It was assumed in this model that drivers exceed the speed limit by not more than 10 km/h. It is assumed that in each time step a vehicle moves at constant speed, so that the distance covered in time $\Delta t$ is $s=v(k)\cdot\Delta t$. It is also assumed that the minimum speed of travel is 10 km/h, which means, roughly speaking, that at that speed vehicles maintain a bumper-to-bumper distance of 5 m.

#### 2.4.2 Travel across intersections

An intersection is a node of the graph that has at least three edges. Nodes with two edges represent places where the road forms a bend, and serve only to provide a graphical representation of the course of the road – they are not places where two streets cross, they do not cause a reduction in traffic speed, and hence they do not increase the time of travel. In the model, the time of travel across an intersection is computed differently for intersections with and without traffic lights.

**Intersections without lights**

Nodes without traffic lights are generally nodes with three neighbours; that is, in most cases, nodes where there is a change in the number of lanes. These are not treated as intersections in the model. As mentioned earlier, local roads, where intersections without lights are encountered, are omitted from the model since they are not significant for traffic flow in the city as a whole. It is assumed in the model that the loss of time when passing such an intersection, usually of roads with equal priority, is 2 seconds. In view of the small number of such
intersections, the size of this delay has little effect on traffic parameters on the scale of the whole city.

Light-controlled intersections
In microscopic models, the conditions of traffic flow on these intersections have been the subject of highly detailed analyses [27-30]. In the present macroscopic model, a simplified, schematic model of a light-controlled intersection is used. The graph of Wrocław’s streets contains 570 such intersections. The same signalling cycle is assumed for all intersections in the city, with a duration of $t_c = 90$ s, where each direction has green lights for a time of $t_g = 30$ s. Similar values were assumed in the studies [31, 32]. The remaining 30 seconds are used for right and left turns, amber signals and other losses. Assuming that vehicles maintain a safe distance of 2 s, this means that the number of vehicles that may cross the intersection on one lane in one direction during one cycle is $c=15$. Given the number of vehicles $n$ waiting to cross the intersection, one may compute the waiting time for a vehicle at the intersection as follows:

$$l_c = n \text{ div } (l \cdot c)$$

(2)

$$n_i = n - l_c \cdot l \cdot c$$

(3)

$$t = l_c \cdot t_c + \frac{n_i}{l_c} \cdot t_g$$

(4)

where:

- $n$ is the number of vehicles waiting at the intersection;
- $c = 15$ is the number of vehicles that can cross in one lane during one signalling cycle;
- $l$ is the number of traffic lanes;
- $l_c$ is the number of full cycles for which a vehicle must wait;
- $n_i$ is the number of vehicles remaining at the intersection after $l_c$ cycles;
- $t_c = 90$ s is the duration of a full cycle;
- $t_g = 30$ s is the time a green light is displayed;
- $l$ is the total waiting time at the intersection.

It is further assumed that the maximum waiting time, irrespective of the length of the queue of waiting vehicles, cannot exceed 180 s, and that the minimum time lost at such an intersection is 3 s.

2.5 Journey route
In the model, the fastest route for each vehicle is determined using Dijkstra’s algorithm [32] for finding the shortest path in a graph. This optimum path is determined on the basis of the traffic density at the journey start time. As the vehicle moves across the graph, there is variation in the traffic density, and thus also in the speed and times of travel, and account is taken of this fact in the model when the journey times between nodes are computed. That is, the time taken to travel from node S to node F is computed from the formula (1), where the density of traffic on the section S-F is taken for the current time at which the vehicle begins travelling along that section.

Dijkstra’s algorithm may be used for any graph, provided that the length of every edge from node to node (the time of travel, in this case) is positive. The graph of city streets satisfies this condition. It nonetheless represents a specific type of graph: each node is linked to a minimum of 2 and a maximum of 6 neighbours. Simulation calculations performed on various graphs suggest that in cases such as this, the computational complexity is proportional to the product of the number of vertices and the mean number of edges leading out of a vertex.

2.6 Simulations
Road traffic simulations based on the model described above are carried out by introducing a set number of vehicles into the road network at each time step. A vehicle has a predefined route, and its position is tracked until it reaches the finish point of its journey. When a vehicle ends its journey, another vehicle is introduced into the street graph, so as to maintain a constant number of travelling vehicles. The computer program performing the simulation illustrates the movement of each vehicle on a map of the city. For the Wrocław street graph, with the number of vehicles set at 15,000 and the time step size at 1 s, the computation time for one hour’s traffic flow is approximately 2 minutes. In the course of the calculations the program records the following parameters for each vehicle: journey time, total waiting time at intersections, time travelling at maximum speed, and length of route covered. At the same time, the model records the traffic loads on individual street sections and intersections. After completion of the simulation the results are archived, and may be analysed graphically and statistically [34].

3 Parameter identification
The model’s main parameter, having a fundamental impact on the results of the computations, is the total number of vehicles making journeys in the city at a given time of day. The city monitoring systems are not currently capable of recording all vehicles on the roads. For this reason, the total number of travelling vehicles was determined through identification of model parameters. Use was made of available data on hourly traffic volumes in Wrocław recorded since 2015 by the Intelligent Transport System (ITS) at several dozen of the largest intersections (Fig. 3).
Validation of the parameters of the model was performed using data from 2015 for the six largest intersections, marked in Fig. 3. Four of them are located on the edges of the city’s central district, and two within that district. The number of vehicles $v_{veh\text{min}}(h)$ travelling in the city at time $h$ was determined by minimising the difference between the real number of vehicles crossing the selected intersections (as observed by ITS cameras) and the number of vehicles returned by the model [35]:

$$\delta(h) = \min_{v_{veh}} \sum_{i=1}^{6} |l_{ITS}(i,h) - l_{Mod}(i,h,\text{veh})|$$  \hspace{1cm} (5)

where:
- $i$ is the intersection number,
- $h$ is the hour of the day,
- $v_{veh}$ is the number of vehicles travelling in the city,
- $l_{ITS}(i,h)$ is the average number of vehicles observed by ITS cameras at intersection $i$ in hour $h$ during 2015,
- $l_{Mod}(i,h,\text{veh})$ is the number of vehicles crossing intersection $i$ in hour $h$ as computed by the model for $v_{veh}$ vehicles travelling in the city.

Identification of the number of vehicles travelling was performed separately for each hour of the day (Fig. 4).

The relative error $RE$ for each hour was computed as the integral relative error:

$$RE(h) = 100\% \frac{\sum_{i=1}^{6} |l_{ITS}(i,h) - l_{Mod}(i,h,v_{veh})|}{\sum_{i=1}^{6} l_{ITS}(i,h)}$$  \hspace{1cm} (6)

The identified number of vehicles travelling in the city is around 15,000 from early morning to late afternoon, and falls to below 1,000 at night. It should be noted that the number of 15,000 is the number of vehicles currently making journeys at every second of the simulation for peak hours. Given that the mean journey time at peak hours is 23 minutes, this means that in the course of one hour during that period, approximately 40,000 vehicles make journeys in the city. The relative error of identification at night-time reaches 20%. This increased error is primarily a result of the fact that the number of vehicles falls to below 1,000. Moreover, at many intersections the traffic lights are switched off at night, and this fact is not reflected in the model, as it would represent a further random or hard-to-determine model parameter. It may also be assumed that at night-time drivers are more likely to exceed the speed limit, and do not always choose a route that minimises the journey time. In the hours between 5.00 and 22.00, however – the times for which traffic modelling has the greatest practical significance – the relative error of validation of the model falls below 10%. Given the large number of parameters of a random nature included in the model, this may be regarded as a sufficient degree of accuracy.

### 4 Verification of the model

The process of verification of a traffic flow model may be carried out in many different ways, depending on the extensiveness and degree of detail of the model [36, 37]. In this study, verification was performed using data from 2016 on traffic volumes at the same intersections that had been used for parameter identification. Traffic volumes were computed for each hour separately, by introducing into the street network the number of vehicles determined at the parameter identification stage. Next, errors were calculated as the difference between the number of vehicles computed using the model and the number recorded by ITS cameras in 2016, using the formula (4). The results of verification are illustrated in Figs. 5 and 6.
At peak hours the relative error is around 15%, while at night-time it exceeds 30%. The relatively large error for night-time may be explained by the same factors as were noted in the analysis of the error for parameter identification. At peak hours the mean speed of vehicles computed by the model for the city as a whole lies in the range 25–26 km/h, which is in agreement with the speeds given on the basis of measurements of actual traffic in spring 2017 by the website www.korkowo.pl [42]. This agreement may be interpreted as additional confirmation of the correctness of the model.

Fig. 6 shows that the numbers of vehicles at intersections computed using the model are generally higher than the empirical values. This results indirectly from the fact that the model was calibrated using data from 2015 and verified using data from 2016. Direct comparison of numbers of vehicles making journeys in 2015 and 2016 shows that in 2016 the number of vehicles fell by approximately 0.6% for the city as a whole, and by as much as 7.2% at the six main intersections selected for use in identification and verification. Nonetheless, based only on a comparison of measurements from two consecutive years, it is difficult to forecast whether the downward trend in the number of journeying vehicles will persist.

Fig. 6. Comparison of modelled and empirical vehicle numbers in 2016.

5 Conclusions

The model presented here for the simulation of road traffic on a macroscopic scale was calibrated using real data supplied by the Intelligent Traffic System in Wroclaw. The relative error of the fit between the modelled and real numbers of vehicles – around 10% for times between 6.00 and 22.00 – was considered to be highly satisfactory from a macroscopic standpoint. For night-time the relative error reaches 20%. The model with parameters determined using data from 2015 was verified using data on traffic volumes in 2016. The relative error of verification for times between 6.00 and 20.00 did not exceed 20%; given the variable nature of traffic volumes in the city, this was considered to be a satisfactory value. Determination of the number of travelling vehicles for each hour of the day separately makes it possible to perform multi-variant simulation calculations for various road incidents for each time of day separately or in a continuous manner for a whole 24-hour period. The traffic model described and verified in this study may be used in analysis of the impact of modifications to the street network, or of possible random events such as accidents and system malfunctions, on traffic parameters [38, 39]. It may also be used to investigate the effect of random parameters of the model, such as distance between vehicles, driver temperament, and duration of traffic light cycles at intersections, on the computed traffic parameters: journey time, distance covered and average speed. Another example of a use of the model is the estimation of the shortening of journey times due to the modification of routes based on time-variable traffic density information, results of which appear in [21]. The traffic flow model may also serve as a constituent of larger models used to study the adverse impact of road traffic on the environment, in particular on atmospheric pollution [40], or to map zones at risk of noise pollution.

References

24. R. Wiedemann, Simulation des Straßenverkehrsflusses (Schrifteneinreihe des IfV, 8, Institut für Verkehrswesen, Universität Karlsruhe, 1974)