A MESOSCOPIC MODEL FOR EVALUATING PERFORMANCE OF SIGNALISED INTERSECTIONS

G. Bernetti, R. Camus, G. Longo
Department of Civil Engineering
University of Trieste
Piazzale Europa, 1 – 34127 Trieste

1. INTRODUCTION

The capacity of a street is related primarily to the geometric characteristics of the facility and the level of service (LOS) depends on intensity and composition of flows. In the case of signalised intersections, an additional element is introduced into the concepts of capacity: time allocation. Time is allocated by traffic signals among conflicting traffic movements that seek use of the same physical space. The way in which time is allocated has a significant impact on performance of the intersection and its approaches.

Performance could be mainly estimated through the computation of queues and delays suffered by vehicles. Detailed methods are developed in the case of simple intersections, but they are not suitable to evaluate performance of multi-junction nodes or small networks. We could use methods developed for extended networks, but they cannot describe completely all the characteristics of this kind of nodes.

A mesoscopic model is presented in order to evaluate performance of all kinds of signal-controlled junctions. Platoons of vehicles are the base elements of the model. Each platoon is identified by few elements and its behaviour depends on the signal plan and on interactions with other platoons. The model uses geometric constructions based on events modifying the system. It can estimate queues and delays suffered by platoons and an aggregate evaluation of performance of each approach and of the whole node is possible.

In the first part we focus on the problem of evaluating delays and level of services by using some well-known methodologies, such as the Highway Capacity Manual model and the TRANSYT model. In the second part the mesoscopic model is described. Finally the model presented is applied to a case study and it is compared with an existing methodology: some elements are pointed out, that make the proposed model more suitable to evaluate delays and levels of service.

2. PERFORMANCE AND LEVEL OF SERVICE OF SIGNALISED INTERSECTION

Performance of signal-controlled junctions could be mainly estimated through the computation of queues and delays suffered by vehicles.

Delay is the difference between the time spent by a vehicle to cross the intersection and the time necessary to do the same in an “ever-green” situation. It is a complex measure and is dependent upon a number of variables, including the quality of progression, the cycle length, the green ratio

allocated to the vehicle in question, the volume/capacity ratio for the lane groups and so on.
Total delay to traffic on an approach is the sum of the delays to all the individual vehicles using the approach to cross the intersection.
Delay is a measure of driver discomfort, fuel consumption, pollution and lost travel time. In particular we are interested in the delay so as to determine the level of service (LOS) of the whole intersection and of any single approach to the node.
Level of service criteria are stated in terms of the average stopped delay per vehicle. Typical criteria are given by the Highway Capacity Manual in Table 1.
LOS A occurs when progression is extremely favorable and most vehicles arrive during the green phase. Most vehicles do not stop at all. LOS B generally occurs with a good progression, even if more vehicles stop than with LOS A, causing higher levels of average delay. At LOS C the number of vehicles stopping is significant, though many still pass through the intersection without stopping. Average delay increases until LOS F: this level is considered to be unacceptable to most drivers.

LOS	AVERAGE DELAY PER VEHICLE [sec]
A	= 10,0
B	> 10,0 e = 20,0
C	> 20,0 e = 35,0
D	> 35,0 e = 55,0
E	> 55,0 e = 80,0
F	> 80,0

Table 1: LOS criteria (Highway Capacity Manual).

We could divide methodologies used to evaluate performance of signal-controlled junctions according to two kinds of nodes: simple intersections and multi-junction nodes.
Note that these values of LOS are defined for simple intersections. In the case of junctions where several simple intersections are involved, LOS limits could be redefined. However for the purposes of this paper we will refer to Table 1.

2.1 Simple intersections (HCM model)

In this case vehicles have to cross just one stopline when facing the intersection. A method to evaluate performance of the node is described in the HCM. The methodology is called “operational analysis” and it results in the determination of capacity and level of service for each lane group as well as the level of service for the intersection as a whole.
Operational analysis is divided into five modules. The first one provides all necessary data (geometric, traffic and signalization conditions). In the second module demand volumes are stated and in the third module saturation flow rate is computed for each of the lane groups established for analysis. The saturation flow rate is based upon adjustment of an ideal saturation flow rate to reflect a variety of prevailing conditions.
In the capacity analysis module volumes and saturation flow rates are manipulated to compute the capacity for each lane group. Finally in the LOS

module delay is estimated for each lane group: delay measures are aggregated for approaches and for the intersection as a whole, and levels of service are determined according to Table 1.
It is meaningful to underline that the HCM methodology is very detailed but it is applicable just in the case of simple intersections.

2.2 Multi-junction nodes (TRANSYT model)

In many cases we have to analyse a spatially limited area with several junctions lying so close to each other that they interact in a much stronger way than in the case of a wide area, both because the conflict pattern is more complicated than in usual intersections and because distances are very seldom sufficient for accumulating queues of any length without crashing the whole system. We call this area a multi-junction node (MJN).
In a MJN we can distinguish three kinds of signals: entry signal (the first one for approaching vehicles) exit signal (the last one for leaving vehicles) internal signal (any other signal). In signalised MJN at least one exit signal is not an entry one. That means that at least one vehicle crosses at least two signals.
The performance of MJN could not be evaluated by using the HCM model, which cannot describe all the features of these nodes.
Since multi-junction nodes hold intermediate features between whole networks and single intersections, the problem of evaluating delays for MJN could be faced with the well known methods developed for networks. These approaches, such as the TRANSYT model, use a graph representation of the network and an iterative process to search the minimum of the objective function.
The TRANSYT model evaluates the average queue over the cycle. For links on which the arrival traffic flow does not exceed the capacity (i.e. the degree of saturation is less than 100 per cent), this average queue corresponds to the rate at which delay is incurred with an identical pattern of traffic arrivals during every cycle and is called the “uniform delay”. Additional delay terms used by TRANSYT are the “oversaturation delay”, due to the steady increase in queues on oversaturated links, and the “random delay”, due to variations in traffic arrivals from cycle to cycle. Oversaturation delay only affects approaches to the MJN and the random delay could be calculated as the TRANSYT formula does. Therefore the model presented in the paper is compared to uniform delay only.
However TRANSYT model shows some limits due to the representation of the signal plan and of its optimization technique. In fact it does not preserve information about the origin and the destination of platoons of vehicles and it is not able to solve some kinds of problems.
For instance, Figure 1 shows a situation that we could call “the double Y problem”. We suppose that the white flow (coming from origin O_1) makes for destination D_2, meanwhile the grey flow (coming from origin O_2) goes to destination D_1. At signal B flows are mixed together: in order to evaluate effective delays suffered by vehicles, a differentiation of platoons is the crucial point of the question.
Normally in TRANSYT no distinction is made between the various types of vehicles which make up the queue on a link. A facility known as a “shared stopline” allows vehicles types to be distinguished within a common queue.

The shared stopline facility allows up to five separate classes of vehicles to be represented in any one queueing situation where, in reality, the classes of vehicles are mixed together.
This facility could be used just for a limited number of flows and for no more than two o three successive signals. Therefore we could say that in general TRANSYT does not resolve the double Y problem.

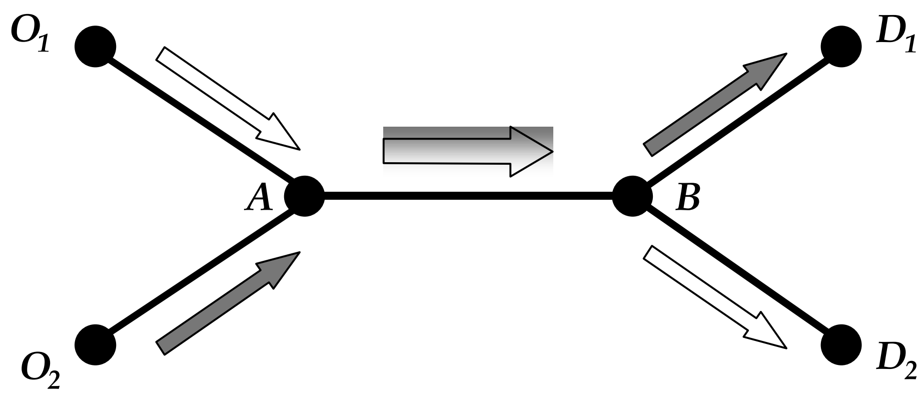

Figure 1: Double Y problem.

As a result TRANSYT can evaluate the total delay and the performance of each signal in terms of delay and queues, but it makes some errors in flow distribution and it cannot evaluate delays of each O/D couple and the aggregated level of service of each approach to the node and of the node as a whole.

3. THE MODEL

We propose a method to face the problem of evaluating L.O.S. of signalised intersection; the method is also suitable in the case of MJN, where common methods are not exhaustive. The method is based on the estimation of delays and queues for each flow, moving from an origin to a destination. The analysis is given through a mesoscopic model: flows are described in terms of platoons of vehicles moving on the network. Each platoon is identified by few elements and its behaviour depends on the signal plan and on interactions with other platoons.
The algorithm does not take into consideration any smoothing (dispersion) process of platoons through links of the node. This assumption does not affect the elementary intersection analysis and holds in MJN or in small networks because of the reduced distances involved.

3.1 Input and Notations

The method just needs classic data used to evaluate L.O.S. in the case of simple intersections and paths followed by flows. Thus input data are very essential:
C, the cycle length;
K, the set of signals in the MJN;

O Ì *K*, the set of entry signals;
s_k, the saturation flow for signal *k* Ì *K*;
gs_k, the start of effective green time of signal *k* Ì *K*;
ge_k, the end of effective green time of signal *k* Ì *K*;
T, the network matrix, the generic element *t(i,j)* of which identifies the length of the link from stop-line *i* to stop-line *j* (if it exists);
R, the flow distribution matrix, where the generic element *r(i,j)* identifies the flow ratio of approach *i* directed to the stop-line *j*;
V, the speed matrix, the element *v(i,j)* of which identifies the link *(i,j)* average speed (if the link exists).
Flows are expressed in terms of passenger car unit (pcu).
For a platoon *i* crossing a generic road section *k* we can also define:
$ps_i(k)$ instant when the first vehicle reaches the section;
$pe_i(k)$ instant when the last vehicle reaches the section;
$qe_i(k)$ instant when the last vehicle crosses the section;
$f_i(k)$ platoon flow assumed uniform during the period (ps_i, pe_i).

3.2 Platoons

The platoon is assumed as the basic traffic unit. Platoons move through the intersection without dispersion. In a graphic representation (Figure 2) we get it with a bar lying on the temporal axis between ps_i and pe_i instants. Bar thickness f_i is proportional to platoon flow.

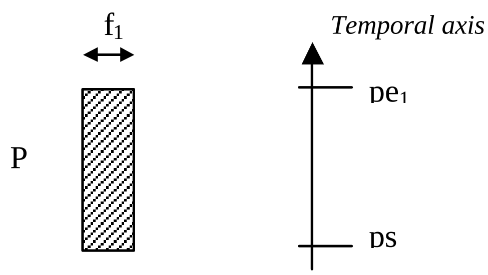

Figure 2: platoon representation.

3.3 Signal Analysis

Behaviour of platoons is analysed by means of the analysis of each signal. We divide the algorithm into three parts: arriving platoons processing, queue analysis and leaving platoons processing.
When platoons coming from distinct origins arrive at signals, they could overlap. This is the case of a confluence, where movements are not conflicting. If we denote with *1* and *2* the arriving platoons (with $ps_1<ps_2$), model develops the following new platoons:

platoon *A* and platoon *B*:

$$\begin{cases} ps_A = ps_1 \\ pe_A = ps_2 \\ f_A = f_1 \end{cases} \qquad \begin{cases} ps_B = ps_2 \\ pe_B = \min(pe_1, pe_2) \\ f_B = f_1 + f_2 \end{cases}$$

platoon *C*:

$$\text{if} \quad pe_1 \le pe_2 \begin{cases} ps_C = pe_1 \\ pe_C = pe_2 \\ f_C = f_2 \end{cases} \qquad \text{otherwise} \begin{cases} ps_C = pe_2 \\ pe_C = pe_1 \\ f_C = f_1 \end{cases}$$

Two sub-platoons form platoon *B*: model preserves information on sub-platoon flows. We define platoon *B* "mixed platoon" (Figure 3).

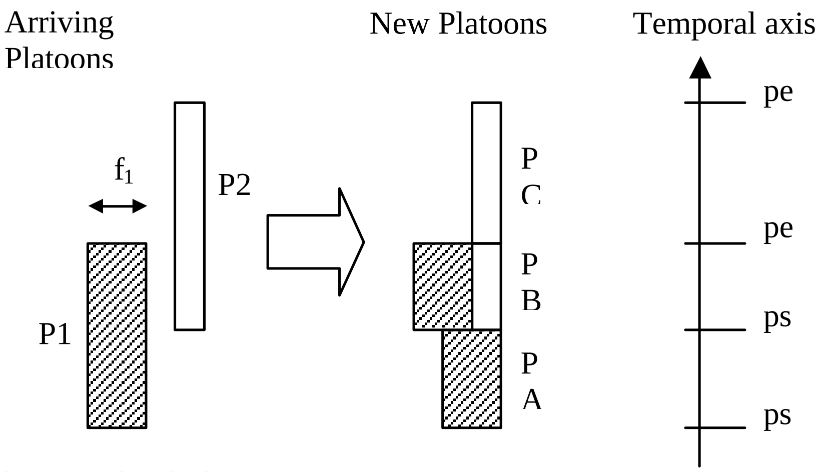

Figure 3: formation of mixed platoons.

We proceed at the same way in the case of more than two overlapping.
If for an arriving platoon *i* at signal *k* is ge_k *Î [ps_i, pe_i]*, then the model divides platoon *i* into two new platoons *A* and *B*:

$$\begin{cases} ps_A = ps_i \\ pe_A = ge_k \\ f_A = f_i \end{cases} \qquad \begin{cases} ps_B = ge_k \\ pe_B = pe_i \\ f_B = f_i \end{cases}$$

Thanks to this elaboration we can then arrange platoons in temporal order over the cycle length.
The behaviour of platoons at signal *k* is analyzed by means of the instants when one of the following events occur:

- gs_k
- ge_k
- ps_i of each arriving platoon *i*
- pe_i of each arriving platoon *i*
- qe_i when the last vehicle of arriving platoon *i* crosses the stop-line.

Cycle is so divided into intervals defined by successive events.
Queue profiles are obtained over the cycle length through elementary geometric constructions. We define the time function $dp_i(t)$ related to platoon *i* and the time function $ds_i(t)$ related to saturation conditions of signal *k* and to platoon *i-1* (when *i>1*). The two functions are defined over the cycle length:

$$dp_i(t) = \begin{cases} 0 & \text{when } t < ps_i \\ f_i \cdot (t - ps_i) & \text{when } ps_i \le t \le pe_i \\ f_i\,(pe_i - ps_i) & \text{when } t > pe_i \end{cases}$$

$$ds_i(t) = \begin{cases} 0 & \text{when } t < \max(ps_i, gs_k, qe_{i-1}) \\ s_k \cdot [t - \max(ps_i, gs_k, qe_{i-1})] & \text{when } \max(ps_i, gs_k, qe_{i-1}) \le t \le ge_k \end{cases}$$

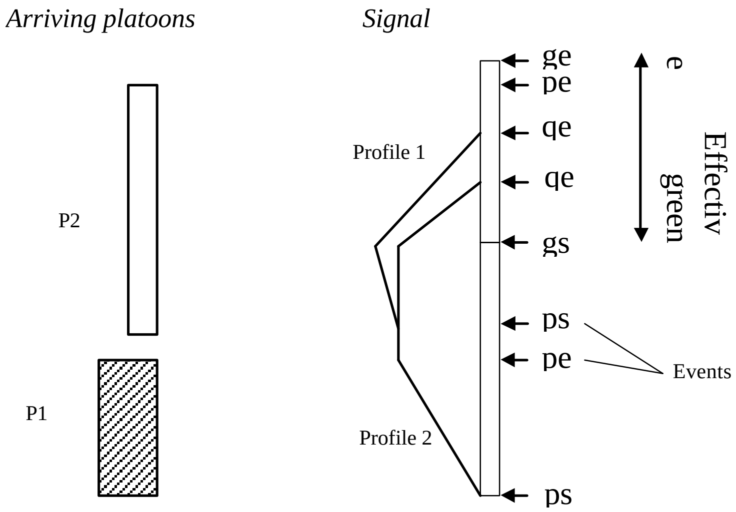

Figure 4: queue profiles.

The queue length q_i of platoon *i* is so defined:

$q_i(t) = \max\{0, y_i(t)\}$ with $y_i(t) = dp_i(t) - ds_i(t)$

Then we compute the end of queue qe_i ($qe_0 = 0$) of platoon *i*:

$qe_i = t_{x-1}+(t_x-t_{x-1}) \times ?_i(t_{x-1})/(?_i(t_{x-1})- ?_i(t_x))$

where:

t_x is the first event when function $?_i$ is negative;

t_{x-1} is the previous event.

Of course the total queue at signal *k* is the sum of single queues. An example is shown in Figure 4.

The delay suffered by each platoon is computed as the area subtended by the queue profile.

According to the order of arrival and queue profiles we now define leaving platoons. In particular if $qe_i > ps_i$ platoon *i* changes shape when it leaves the stop-line. If $qe_i = pe_i$ then leaving platoon has the following features:

$$\begin{cases} ps_i = \max(ps_i, qe_{i-1}) \quad \text{with } q_0 = gs_k \\ pe_i = qe_i \\ f_i = s_k \end{cases}$$

If $q_i < pe_i$ vehicles composing a platoon are subjected to different events: a part of the platoon is forced to wait in queue, while some vehicles can cross the stop-line without stopping. The platoon itself is divided into two new platoons *i1* and *i2*. Platoon *i1* has the same parameters of previous case, while platoon *i2* is so defined:

$$\begin{cases} ps_{i2} = qe_i \\ pe_{i2} = pe_i \\ f_{i2} = f_i \end{cases}$$

Figure 5 shows an example.
Leaving platoons preserve information about suffered delay.

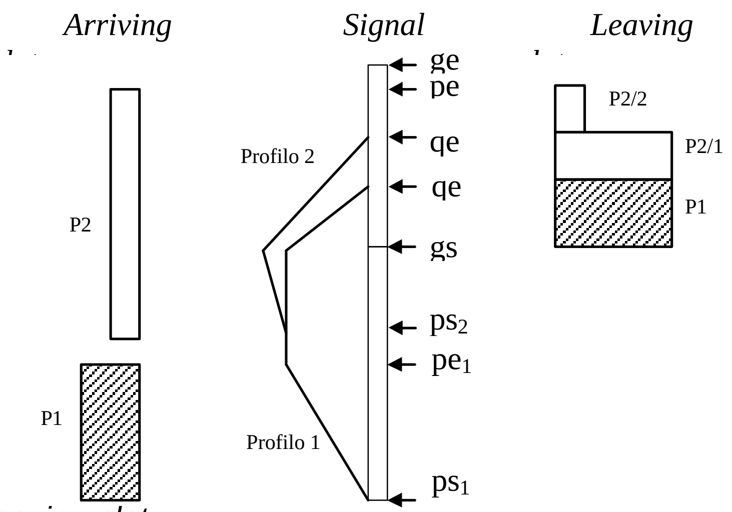

Figure 5: leaving platoons.

3.4 Node Analysis

Depending on the *T* matrix of the network, platoons move across the MJN. Leaving a signal platoons spread according to the *R* matrix and to the *V* matrix of speeds.
Note that signal analysis is complete if all the arriving platoons are already defined. It means that signals are analysed following a sequence depending on MJN topology.
As a result, for each stop-line the algorithm allows the evaluation of delay suffered by each platoon. Moreover for each platoon leaving the MJN it is possible to recognise delays suffered at each encountered signal.

4. APPLICATION AND COMPARISON: A CASE STUDY

In order to get a validation of the model proposed, it has been applied to several intersections, located in different city in Italy, and we made a comparison with consolidated models.
In particular The TRANSYT model is applied to the same nodes and results are compared. Thanks to this comparison we can underline advantages offered by the model presented and some limits of TRANSYT come out: in fact the TRANSYT model cannot give a correct evaluation of LOS and some errors due to the TRANSYT structure arise. Such errors are mainly related to the loss of information about origin and destination of platoons along the node.
As an example we apply now the model to an actual MJN (Figure 6): we consider the case of Piazza Maggi in Milan (Italy).

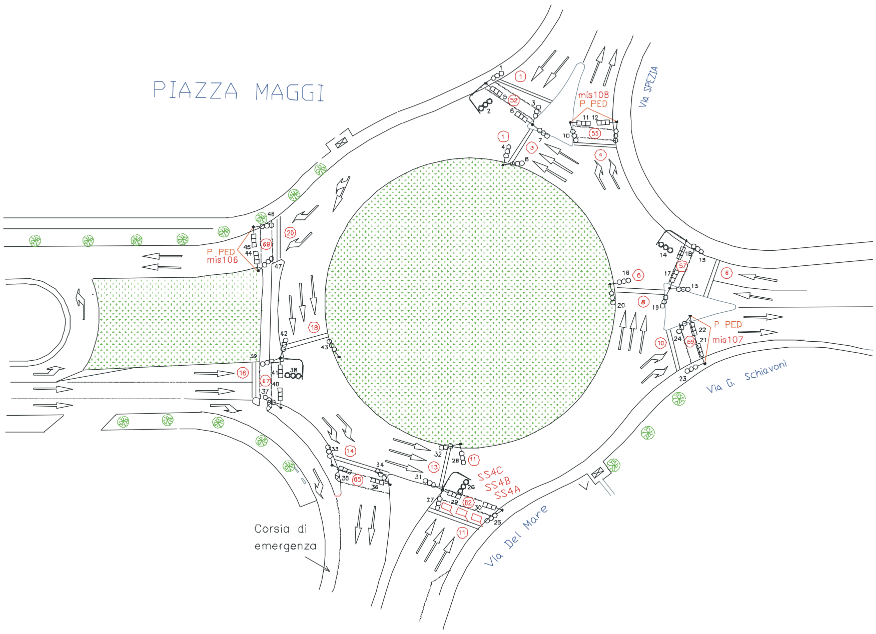

Figure 6: Piazza Maggi (Milan) – layout of intersection.

Cycle length is equal to 90 sec. Signal plan, flows and flow distribution are given in Table 2: for each signal the Table gives Start and End green instant, approaching flow and saturation flow; the flow distribution matrix R and the network matrix T are also reproduced. It is assumed that average speed is 36 km/h over the whole network.

The same data are used as input for the TRANSYT program. TRANSYT uses a formula of dispersion of platoons through links, but it does not affect the comparison (differences are less than 3%). Anyway no platoon dispersion is forced in TRANSYT input data.

In Table 3 results of the two models are compared. For each signal, delays estimated by the model presented are expressed in terms of sec/PCU (passenger car unit) and they are compared with uniform delays given by TRANSYT.

Differences are not significant when we consider entry signals (signals 1, 16, 11 and 6): little variations are due to TRANSYT output and time resolution. In fact in the TRANSYT algorithm the cycle time of the signals is divided into a number of equal intervals ("steps"); these are typically 1 to 3 seconds long. All TRANSYT's calculations are made on the basis of the average values of traffic flow rates which are expected to occur during each step. This introduces some errors, especially at starting and ending of effective green periods. Other errors occur because of the TRANSYT output: it gives no more than two significant digits.

On the contrary, differences could be considerable when analysing internal signals. The TRANSYT model presents certain limits in the analysis because TRANSYT does not identify the origin of flows: as a result, when vehicles cross signal 8, signal 8 feeds signal 3 and it spreads all the flows, apart from their origin. At signal 8, some vehicles coming from signal 11 are directed to signal 3 (see flow distribution matrix), meanwhile no vehicles coming from signal 16 feed signal 3.. The model proposed is able to take into consideration this fact, because it could distinguish origins of platoons. The different result between the models is so justified.

SIGNAL PLAN AND FLOWS								
Signal Number	Start Green [s]	End Green [s]	Approaching Flow [pcu/h]	Saturation Flow [pcu/h]				
1	54	82	900	3600				
3	86	50	0	5400				
4	16	86	0	3600				
6	84	44	1800	3600				
8	48	80	0	5400				
10	73	52	0	3600				
11	51	76	900	3600				
13	80	47	0	5400				
14	34	1	0	3600				
16	3	33	1100	3600				
18	37	89	0	5400				
20	80	59	0	3600				
NETWORK MATRIX T [m]								
Signal	**3**	**4**	**8**	**10**	**13**	**14**	**18**	**20**
1	0	0	0	0	0	0	92	78
3	0	0	0	0	0	0	92	78
6	57	47	0	0	0	0	0	0
8	57	47	0	0	0	0	0	0
11	0	0	82	72	0	0	0	0
13	0	0	82	72	0	0	0	0
16	0	0	0	0	52	42	0	0
18	0	0	0	0	52	42	0	0
FLOW DISTRIBUTION MATRIX R [%]								
Signal	**3**	**4**	**8**	**10**	**13**	**14**	**18**	**20**
1	0	0	0	20	20	40	60	40
6	80	20	0	0	0	30	30	50
11	30	40	70	30	0	5	5	25
16	0	40	40	60	100	0	0	0

Table 2: input data (Piazza Maggi).

	MODEL					TRANSYT		DIFFERENCES	
Signal number	*Delay/C [sec]*	*Delay/h [sec]*	*Veic/C*	*Veic/h*	*Delay sec/PCU*	*Delay PCU-h/h*	*Delay sec/PCU*	*[sec]*	*[%]*
1	641	25627	22	900	28,47	7,1	28,40	-0,07	0
3	99	3951	43	1710	2,31	1,6	3,37	1,06	46
4	58	2308	29	1160	1,99	0,5	1,55	-0,44	-22
6	800	32000	45	1800	17,78	8,9	17,80	0,02	0
8	313	12504	27	1070	11,69	3,6	12,11	0,43	4
10	48	1929	28	1110	1,74	0,6	1,95	0,21	12
11	704	28167	22	900	31,30	7,8	31,20	-0,10	0
13	16	621	32	1280	0,49	0,4	1,12	0,64	132
14	34	1373	24	945	1,45	0,2	0,76	-0,69	-48
16	792	31680	28	1100	28,80	8,8	28,80	0,00	0
18	204	8175	28	1125	7,27	2,6	8,32	1,05	14
20	106	4230	37	1485	2,85	1,2	2,91	0,06	2

Table 3 – Comparison between the model proposed and TRANSYT results.

We could highlight the problem by using a graphic representation. In Figure 8 a screenshot of the software developed using the presented model is shown. Two windows describe a simulation of signal 8 (on the left) and signal 3 (on the right). The upper part of each window define arriving platoon, then queue development is illustrated and finally leaving platoons are specified. Platoons are identified by using different colour, depending on the origin.
As we can see, just platoons with a light colour (coming from signal 11) are directed from signal 8 to signal 3. Figure 7 describe TRANSYT output referred to the same signals 8 and 3. We can observe that a fraction of all the leaving platoons of signal 8 (identified by the outline) is directed to signal 3. It could be underline by the shape of the outlines of leaving vehicles of signal 8 and arriving vehicles of signal 3. The distribution process is very different in the two models.
This simulation error of TRANSYT could be avoided by using the “shared stopline” facility, as described above. This is just possible in simple cases and for a limited number of flows.

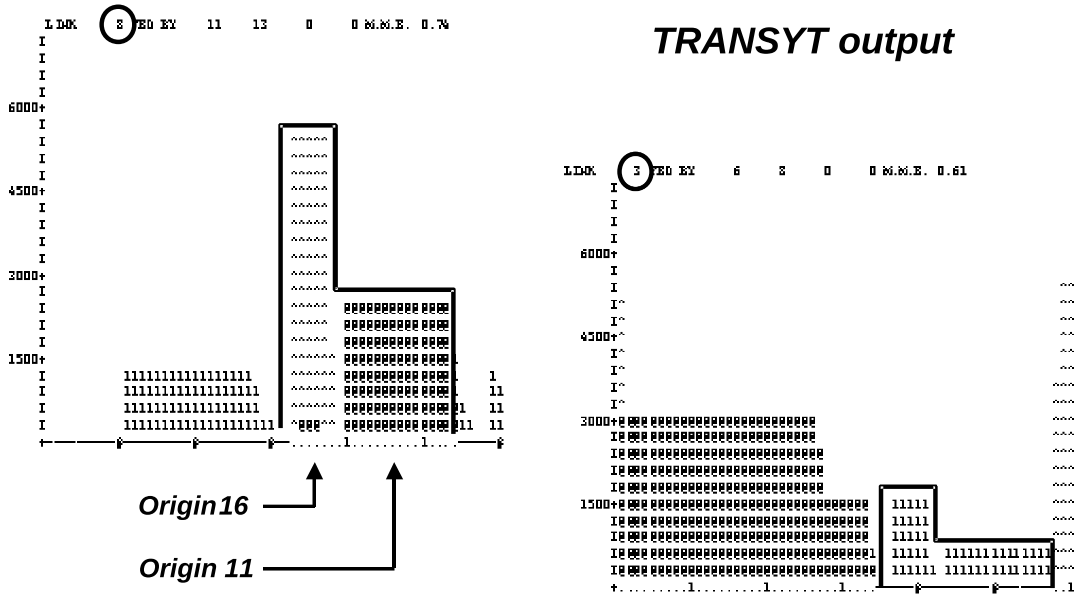

Figure 7: TRANSYT errors.

In any case TRANSYT output cannot give us other information. As we already stressed TRANSYT does not allow determination of a platoon’s “history”: all the platoons approaching a signal are mixed together into the same histogram. Thus it is impossible to evaluate delay suffered by vehicles divided by O/D. We can try to do that by adding delays suffered at each signal of the considered path, but the outcome is not good enough and it is conceptual wrong, too. So it is impossible to estimate the effective LOS of the network.
On the contrary the model presented let us evaluate delay for each O/D couple. Thus we can determine LOS for each O/D path.
We can also calculate the weighted-average delay per vehicle for each approach O: it is found by adding the product of the O/D platoon flow and the O/D delay for each platoon on the approach O and dividing the sum by the total approach flow rate.

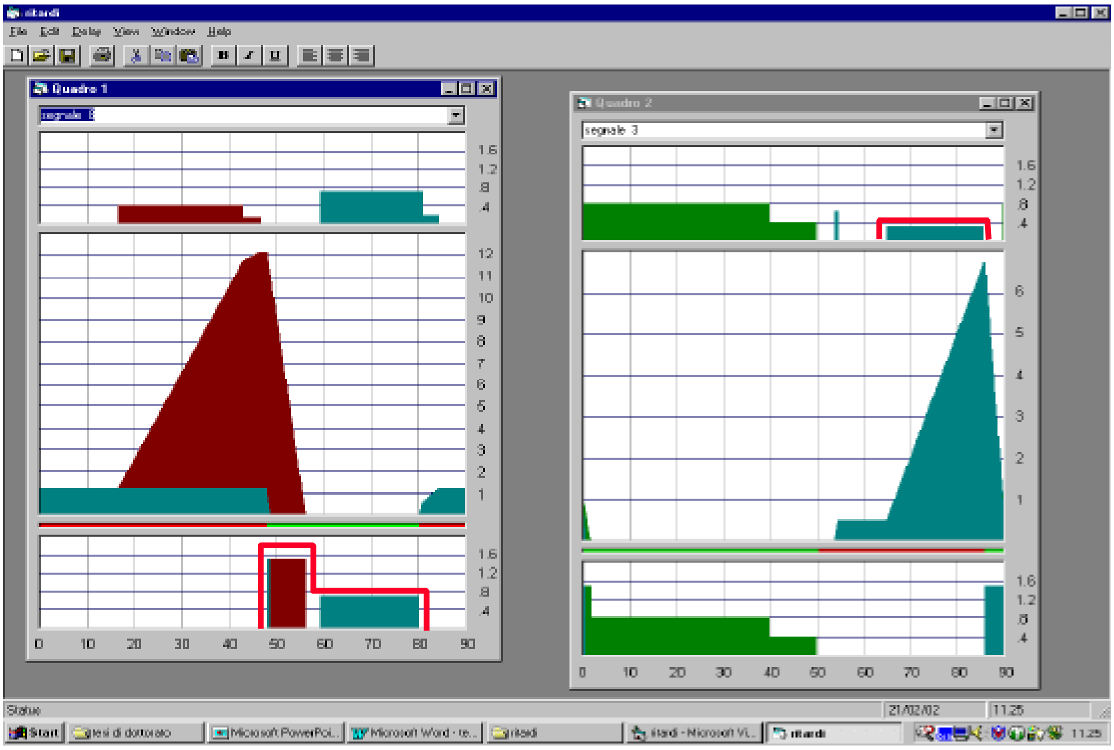

Figure 8: Output of the model.

DELAYS AND LOS FOR EACH O/D COUPLE					
Origin	Destination	Delay [sec]	PCU	Delay/PCU	LOS
1	10	149	5	33	C
1	14	271	9	30	C
1	20	357	9	39	D
6	4	197	9	21	C
6	14	426	14	31	C
6	20	402	23	17	B
11	4	320	9	35	C
11	10	258	7	38	D
11	14	91	1	80	E
11	20	285	6	50	D
16	4	582	11	52	D
16	10	475	17	28	C

DELAYS AND LOS FOR EACH APPROACH				
Approach	Delay [sec]	PCU	Delay/PCU	LOS
1	778	23	34	C
6	1025	45	22	C
11	954	23	42	D
16	1057	28	38	D

DELAYS AND LOS FOR THE INTERSECTION			
Delay [sec]	PCU	Delay/PCU	LOS
3814	118	32	C

Table 4 – LOS evaluation.

In the same way we calculate the weighted-average delay per vehicle for the MJN as a whole: it is found by adding the product of the approach flow rate and the approach delay for all approaches and dividing the sum by the total MJN flow rate.
Thus we evaluate LOS for each approach as shown in Table 4. Finally LOS for the MJN is estimated.
Note that we have adopted limits specified in Table 1, even if Table 1 criteria are calibrated for simple intersections only and they may be unsuitable in the case of MJN or little networks.

5. CONCLUSIONS

Signalised nodes are often the critical points of networks: getting good control strategies for them is a necessary condition for a successful solution of the traffic problem. Due to the possible high values of delay, if a saving of only a few per cent could be obtained by using improved methods to set signals, the time saving would be considerable.
The problem of evaluating the performance of signalised intersections is faced. Existing models are often inadequate, especially when we consider multi-junction nodes or small signalised network. We present a mesoscopic model in order to estimate delays and levels of service for all kinds of signalised intersections. The model is applied to a case study and the comparison with a consolidated methodology, such as the TRANSYT model, let us to point out some elements, which show that the proposed model is more suitable in evaluation of signal plans.
The model allows an evaluation similar to that obtained with HCM method for simple intersections, but it could be used for all kinds of intersections. The model is able to give an overall indicator of performance of the node and of each approach.
A future development of this method could be the introduction of delay divided by O/D into the objective function of optimisation algorithms. Moreover the method could be compared with a micro-simulator approach, which however needs more detailed data. Finally we could use the model in order to estimate turn penalties in network assignment by representing multi-junction nodes as an equivalent simple intersection.

REFERENCES

Bernetti, G., Camus, R., (2001) L.O.S. evaluation in multi-junction nodes, *Proceedings of the International Conference on Traffic Science 2001,* Ljubljana (Slovenia).

Highway Capacity Manual (1997) Special Report 209, third edition, Transportation Research Board, Washington D.C.

Camus, R., D’Amore, A., Ukovich, W. (1983) Optimal Fixed Time Vehicular Control for Multi-Junction Nodes Using Mathematical Programming, *Proceedings of the 4th IFAC/IFIP/IFORS International Conference on Control in Transportation Systems*, Baden Baden.

Cantarella, G. E., Sforza, A., Viola, F. A. (1998) Metodi di regolazione delle intersezioni semaforizzate, *Modelli e metodi per l'ingegneria del traffico*, FrancoAngeli, Milano.

Cantarella, G. E., Sforza, A., Viola, F. A. (1998) Analisi delle prestazioni di intersezioni semaforizzate, *Modelli e metodi per l'ingegneria del traffico*, FrancoAngeli, Milano.

Webster, F. V., Cobbe, B. M. (1966) *Traffic Signals*, Road Research Technical Paper n. 56, HMSO, London, 1966.