STANDARDIZATION OF TRAFFIC SIMULATION MODELS THROUGH VERIFICATION AND VALIDATION

Ryota Horiguchi, Dr. Engineering
President of i-Transport Lab. Co., Ltd.
2-12-404 Ageba-cho Shinjuku-ku Tokyo 162-0824 Japan,
phone and fax: +81-3-5261-3077, e-mail: horiguchi@i-transportlab.jp

INTRODUCTION

This paper introduces the movement of standardization of traffic simulation models in Japan. We now find many reports of the simulation studies in business scene. However, the users except model developers would have less knowledge about the simulation models, because it is difficult to fully understand the nature of the models only by reading literatures or manuals. The simulation is sometimes criticized as "black box", and reconciles itself to unreliable technique.

In order to cope with this criticism, Simulation Committee in Japan Society of Traffic Engineers (Sim@JSTE) have encouraged model developers to disclose the nature of their models through verification [Horiguchi, et al. 2000] and validation [Horiguchi, et al. 1998] with the purpose to promote the utilization of traffic simulation. The basic ideas of verification and validation are as follows:

“Verification” is a series of simple tests to confirm that fundamental model functions are properly programmed as in the specification. The simulated result is compared with what the result should be obtained from the well-authorized theory. In order to individually examine each of functions and also to get the theoretical solutions to be compared, we should use virtual data on a simplest network.

“Validation” seems quite similar to “Verification” in general. However, we clearly distinguish “Validation” such that it is the evaluation of the model specification using real field data. Even if the model is verified as in the specification, the model specification itself may not be adequate to describe real traffic phenomena. The model cannot be practically applicable, if actual traffic situations are not sufficiently reproduced due to the incomplete model specification. Furthermore, the model performance as a system should also be confirmed, such as whether the execution of the model can be finished within a practical computation time.

One of our outcomes is the verification manual [JSTE. 2001a], which describes the standard verification procedure so as to be applied to various different types of models. The manual contains a series of basic tasks to check the reproducibility of traffic conditions of a model by applying simple but ideal dataset. Each of the basic tasks evaluates 1) vehicle generation pattern, 2) bottleneck capacity and saturation flow rate at an intersection, 3) shockwave propagation, 4) capacity of merging and diverging section, 5) right (left) turn capacity decline at a signalised intersection, and 6) dynamic route choice behaviour. In each task, simulation results are to be compared with 'well-known' theories in traffic engineering. A theory, which
sometimes too much simplifies the traffic phenomena, can give us some good standing point to understand the models’ behaviour. We do not, therefore, require the model completely follows the theory, but when the simulation result is different from the theory, the verifier of the model can explain why the result shows such discrepancy to vindicate the model itself.

Several simulation models that are practically used in Japan have been evaluated based upon the proposed verification process. We have verified seven pilot models, such as AVENUE [AVENUE. WWW site], SOUND [Yoshii, et al. 1995], tiss-NET [Sakamoto, et al. 1998], Paramics [Paramics. WWW site], NETSIM [NETSIM. WWW site], REST [Yoshida, et al. 1999], and SIPA [Yokochi, et al. 1999] along the verification manual.

Other output is the benchmark datasets, which are real field data observed and processed well [Hanabusa, et al. (2001)]. Although verification can help us to comprehend the models' behaviour, it does not tell us the applicability for the real world where various traffic phenomena embedded and affect each other. Therefore, the model developers have to show the evidence that those models can reasonably reproduce such complex situation through the validation with benchmark dataset. The benchmark datasets have been gathered so that developers can utilize them to validate their models, since in general acquisition of real field data is substantially expensive and time consuming. Some of the pilot models were validated through the application with benchmark dataset [Horiguchi, et al. 1996] [Sawa and Yamamoto. 2002].

These outputs can be found in “Clearing House of Traffic Simulation Models” at an Internet website [JSTE. 2001b]. The developers and users of simulation models are also encouraged to publish their experiences on verification and validation through Clearing House. Our activities for verification and validation can make sense as the objection against the "black-box" criticism only when the ability of the model is disclosed.

**STAGES OF MODEL DEVELOPMENT**

In order to clarify the standpoint of verification, let us start with stages of the model development. We consider the following five stages:

**Framework**
According to purposes of model applications, we have to first construct the model framework; that is, what the model input and output are, what kind of traffic phenomena should be described, and so on. Then, we have to conceptually decide vehicle motions as well as travellers’ behaviour to be incorporated so as to reproduce the traffic phenomena concerned.

**Specification**
The model framework defined above should be specified in more detailed way. For instance, we should decide how long the scanning interval is, and what kind of car-following models as well as route choice models should be employed, etc.

**Implementation**
This process consists of programming to run the model contrived in the previous stage on a computer and debugging to check if the computer operates according to the algorithm. Debugging is just the fixing coding errors and it must be distinguished from verification described in the next.

**Verification**
Verification is a series of simple tests to confirm that fundamental model functions are properly
programmed as in the specification. As mentioned earlier, the simulated result is compared with what the result should be obtained from the well-authorized theory. In order to individually examine each of functions and also to get the theoretical solutions to be compared, we should use virtual data on a simplest network.

**Validation**

“Validation” seems quite similar to “Verification” in general. However, in this paper, “Validation” must be clearly distinguished such that it is the evaluation of the model specification using real field data. Even if the model is verified as in the specification, the model specification (or the model framework) itself may not be adequate to describe real traffic phenomena. The model cannot be practically applicable, if actual traffic situations are not sufficiently reproduced due to the incomplete model specification. Furthermore, the model performance as a system should also be confirmed, such as whether the execution of the model can be finished within a practical computing time.

**MANUAL OF STANDARD VERIFICATION PROCESS**

Verification is a sort of virtual test using ideal network and demand configurations to qualify basic phenomena on road traffic. Generally, we do not expect a simulation model to reproduce real traffic situation perfectly, since some simplification in vehicle motions and user behaviours have to be made and also discretization of time and space is required to some degree. Accordingly, in the verification, the simulation result is not expected to exactly agree with the theoretical results. When the simulated result is far different from the theory, we should revise the programming. However, in most of the cases, the simulation would perform slightly different way. In these situations, we should utilize the verification to understand the model characteristics. Establishing this linkage will provide us very helpful information to understand the model characteristics. At the same time, it is considered important to define the relationship between certain model parameters and model behaviour.

**Features to be considered through Simulation**

We have included six basic features in the verification manual, which should be at minimum considered by the network simulation models.

**Generation of Vehicles and Flow Conservation**

For implementation of simulation, it is necessary to generate the traffic at the entry end according to the arrival distribution of vehicles from outside the study area. Most of the simulation models seem to assume random arrival at a network boundary section, but there might be some other arrival patterns to be adopted by considering the objective of the simulation study. For example, the uniform arrival may be assumed in some cases of the analysis for over-saturated traffic conditions, in order to avoid the undesirable tendency of pseudo-random series. The verification process here requires whether the generation pattern assumed in the model really achieved. It should be also checked whether the number of vehicles generated in a certain time period is equal to or different from the given volume.

Once a vehicle generated, it must not disappear until it reaches its destination. Even in the case that vehicle queue spills out of study area, newly generated vehicles are added to the end of the point queue outside the network and will flow into the network after sufficient time period. Simulation models, so that, must keep this flow conservation law not only at every links but also outside of network.

**Bottleneck Capacity / Saturation Flow Rate at Link’s Downstream End**

As the discharging flow rate from a bottleneck section like sags or tunnels contributes to the reproduction accuracy of the delay caused by the congestion at the bottleneck, it is essential that the capacity of the
bottleneck should be reproduced in a stable manner during the simulation.

Even on surface streets in under-saturated conditions, a vehicle may have the delay caused at signalised intersections. The outflow from an intersection continues at the saturation flow rate till a vehicle queue developed during the red vanishes. It is important to clarify how the saturation flow rate is reproduced in the simulation model as for the bottleneck capacity.

**Growing and Shrinking Traffic Jam Consistent with Shock Wave Theory**

When traffic jam beginning at a bottleneck grows to the upstream link, even traffic that does not need to pass through the bottleneck may be also affected. As a difference in the jam’s growing/shrinking speed results in difference in the degree of influence on the total delay upon whole network, it is important to reproduce this phenomenon by using physical-queue to reasonably maintain the traffic density of congestion. The verification of these phenomena is made by comparing the shock wave speed simulated with the one based on the shock wave kinematics, as shown in Figure 1.

![Figure 1: The growing speed of jam is determined by the arrival demand and the bottleneck capacity.](image)

For surface streets, on the other hand, even if a signalised intersection is under-saturated, the vehicle queue grows and shrinks in every cycle. The tail of the queue moves with some time lag at the begging of the green phase due to drivers’ response delay at departure. Because of this time lag, when two signalised intersections are close and the queue heading to the one intersection spreads beyond another, there may be the case that the vehicles in the tail of the queue cannot pass through the near intersection depending on the signal offset. Therefore, the simulation model that is considering signal control effect must reasonably reproduce this phenomenon including shock wave propagation.

**Capacity of Merging and Diverging Section**

Not only sags or tunnels but also merging and diverging sections can be the most remarkable bottleneck of highways. At a congested merging section, the travel time on each approaching branch may vary with the merging ratio even if the capacity of the merging section stays constant. Contrary, the capacity of the diverging section is constrained by the capacities of downstream links and may change depending on the proportion of the demand to each branch. The verification step includes these merging and diverging configurations.

**Gap Acceptance of Right (Left)-Turn at a Signalised Intersection**

In ordinary streets, it is daily observed that vehicles waiting for right (left)-turn in the signalised intersection sometimes obstruct travel of followers and cause congestion. Such vehicles are waiting to find an acceptable gap in the opposing straight-through traffic in the green phase, and consequently the right (left)-turn capacity declines according to the opposing traffic volume. A simulation model that treats a
signalised intersection would be required to describe such relationship between the turning capacity and the opposing traffic volume by some set of gap acceptance parameters.

Drivers' Route Choice Behaviour
Modelling for drivers’ route choice behaviour considered in simulation will be classified as follows:

a) No route choice,
b) Dynamic User Optimal (DUO) principle,
c) Dynamic User Equilibrium (DUE) assignment,
d) Probabilistic route choice.

Of these models, the one using a) above is considered applicable to evaluation of the short-term traffic management that need not consider routes of drivers, or to a network without any alternative routes. Verification of these models is not necessary because it is equivalent to the verification at a merging/diverging section.

On the other hand, the simulation model using principles of b), c) or d) let a driver select an appropriate route according to the presented information for routes. This type of models is frequently used to evaluate the operational policy such as dispersing the traffic spatially by means of informative service or road construction. Models with route choice can be verified using a simplified network, e.g. with two routes for one O-D pair, to avoid the difficulty to figure out the theoretical flow pattern to be compared with the simulation result. It is also interesting to examine results by changing settings of simulation model such as an update interval of route costs and locations where the drivers can receive the travel cost information.

UNDERSTANDING MODELS' NATURE THROUGH VERIFICATION AND VALIDATION

In advance of case studies, a user might think what simulation model can be applicable to the subjective problem. For that, the user must have good knowledge about his/her available models’ nature to choose appropriate one of them. Since the major purpose of this paper is to give a reader some idea how the verification and the validation processes referred in the manual will be helpful to understand the nature of model, let us introduce several results of verification and validation as examples in this section. The details of verification process and further results of verification are referred by [Horiguchi and Kuwahara. 2002].

Generation of Vehicles

For implementation of simulation, it is necessary to generate the traffic at the entry end according to the arrival distribution of vehicles from outside the study area. Most of the simulation models seem to assume random arrival at a network boundary section, but there might be some other arrival patterns to be adopted by considering the objective of the simulation study. For example, the uniform arrival may be assumed in some cases of the analysis for over-saturated traffic conditions, in order to avoid the undesirable tendency of pseudo-random series. The “Standard Verification Process Manual” [JSTE. 2001a] requires whether the generation pattern assumed in the model really achieved.

Adding to this, it should be also checked whether the number of vehicles generated in a certain time period is equal to or different from the given volume. Figure 2 indicate the results with different random seeds for AVENUE [AVENUE. WWW site] and tiss-NET [Sakamoto, et al. 1998], both of which assume random arrival in vehicle generation. AVENUE always generates the same number of vehicles as the given demand level (Q=500, 1000, 2000 [veh./hr]), on the other hand tiss-NET varies its results with each random seed.
The results coming from the difference in the attitude of their ‘specification’ stages can be known only through the qualify tests in verification. It gives meaningful implications that literatures would not tell us. For this case, a user of the simulation model that has the same nature as tiss-NET in vehicle generation should realize that he or she has to repeat the simulation for the same network and demand configuration with different random seeds. The user also has to be careful in choosing the set of random seeds not to be biased in the number of generated vehicles against to the given demand setting. Subsequently, the user must evaluate the variation of the number of generated vehicles for each calculation.

**Figure 2: Total number of generated vehicles – (Left: AVENUE, Right: tiss-NET)**

**Traffic Flow Characteristics and Bottleneck Capacity of C-F type Models**

As the discharging flow rate from a bottleneck section like sags or tunnels contributes to the reproduction accuracy of the delay caused by the congestion at the bottleneck, it is essential that the capacity of the bottleneck should be reproduced in a stable manner during the simulation.

According to the procedure described in the manual, the traffic flow characteristics of each C-F type simulation model must be identified in its verification process. Here, let us introduce the verification of Paramics [Paramics. WWW site] and SIPA [Yokochi, et al. 1999] as examples, both of which have a dozen of model parameters concerning the driving behaviour and the link performance.

**Table 1: Major model parameters of Paramics and SIPA**

<table>
<thead>
<tr>
<th>Model</th>
<th>Driving behaviour</th>
<th>Link performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paramics</td>
<td>Minimum headway,</td>
<td>Headway coefficient,</td>
</tr>
<tr>
<td></td>
<td>Maximum acceleration,</td>
<td>Limit speed,</td>
</tr>
<tr>
<td></td>
<td>Driving aggression, etc.</td>
<td>Gradient, etc.</td>
</tr>
<tr>
<td>SIPA</td>
<td>Target headway,</td>
<td>Allowable minimum headway,</td>
</tr>
<tr>
<td></td>
<td>Target speed,</td>
<td>Limit speed,</td>
</tr>
<tr>
<td></td>
<td>Maximum acceleration,</td>
<td>Gradient, etc.</td>
</tr>
<tr>
<td></td>
<td>Response delay, etc.</td>
<td></td>
</tr>
</tbody>
</table>

The major parameters of each model are listed in Table 1. The meanings of some are clear, e.g. maximum acceleration or limit speed, but not all. For instance, how is the "minimum headway" of Paramics difference from the "target headway" of SIPA, what is the "headway coefficient" of Paramics, or is the inverse of "allowable minimum headway" of SIPA equal to the link capacity? Even if their software manuals or technical papers state the meanings of the parameters, they are mostly conceptual explanations. It is still mysterious how each of the parameters effects on the bottleneck capacity.

Our interest here is to understand the quantitative relationships between the model parameters and the bottleneck capacity reproduced in the simulation. Furthermore, we would like to find the most sensitive parameters through the verification process, because it must be the most efficient strategy to fit the
The simulation result to an actual traffic condition by changing the most sensitive parameters.

The left figure in Figure 3 shows a portion of the results of Paramics. The dots in the figure indicate the volume-density plots observed with varying major parameters. The shape of the dots is associated with the sort of varied parameters. The remarkable point is that the decline of the flow rate is found only when the "headway coefficient" of the bottleneck link is 1.5 (dots surrounded by the circle) and otherwise there are no effects. This implies that only the changes on the "headway coefficient" of the bottleneck link does affect to the bottleneck capacity while others have less influence.

The right figure in Figure 3 also shows the result of SIPA in the case that the "minimum headway" of the bottleneck link changes from 2.0 seconds to 3.0 seconds. Theoretically, the minimum headway must be equal to the inverse of the capacity, so that the bottleneck capacity must be 1200 pcu per hour if the minimum headway is 3.0 seconds. However, the bottleneck capacity reproduced in the verification is slightly greater than the theoretical value. We may realize that the "minimum headway" of SIPA is similar but different parameter from the link capacity.

There are common findings through the verification of bottleneck capacity for the car-following type models:

i) Most of them have the parameters that affect to the minimum headway of each link.
ii) Such parameters have strong influence on the bottleneck capacity but others have less influence.
iii) Such parameters are not exactly equivalent to the inverse of the bottleneck capacity.

There are some implications obtained from i) and ii). Even if we use so-called microscopic simulation models, we have to be rather careful in calibrating the link parameters related to headways than those to driving behaviours. In this sense, such microscopic simulation models are essentially equivalent to the macroscopic simulation models that require the capacities of links.

**Saturation Flow Rate at a Signal Intersection**

Even on surface streets in under-saturated conditions, a vehicle may have the delay caused at signal intersections. The outflow from an intersection continues at the saturation flow rate till a vehicle queue developed during the red vanishes. It is important to clarify how the saturation flow rate is reproduced in the simulation model as for the bottleneck capacity.

For the verification of the saturation flow rate at a signal intersection, the tester is required to show the profile of discharging traffic within a signal cycle. Let us introduce the result of SOUND [SOUND. WWW site], which has the combined flow model: car following for expressways (SOUND/express) and queuing vehicle lists for arterial roads (SOUND/A-21). The former calculates each vehicle speed
according to the spacing-velocity (S-V) function given to each link. The S-V functions can be identified through macroscopic surveys of traffic flows. On the other hand, the latter assumes the point-queue at the downstream end of each link. The point-queue of each link accepts vehicles up to the jam density and discharges them at the saturation flow rate of the link within the green signal.

Figure 4 illustrates the profile of vehicle discharging for SONUD/A-21. As SOUND/A-21 is a sort of "Q-K type" models, the discharging flow rate at saturation is expected to strictly agree with the given saturation flow rate. Now we may confirm from the figure that the simulation result attains the given saturation flow rate as 1600 pcu/G1hr (an effective green hour) in average.

![Figure 4](image)

Figure 4: Link discharging profile at a signal intersection -- SOUND/A-21

There is another point to be discussed in Figure 4. The discharging flow rate of SOUND/A-21 immediately goes up to the saturation flow rate when the signal turns to green. In the actual situation, it takes some time to discharge the flow at the saturation flow rate because of the response delay of drivers. The tester of SOUND gives the reason to this point as follows:

- Instead of neglecting the starting delay, a vehicle cannot flow out during yellow signals in order to adjust the effective green time.
- At normal intersections, the duration of green signal is nearly equal to the effective green time so as to take the yellow interval as much as the starting delays.

**Network Configuration to be Applicable for the Simulation Model**

Not only the verification of simulation models but also the validation can give us useful information concerned with the models’ nature. Figure 5 illustrates the surface street network in Kichijoji-Mitaka area, Tokyo, on which precise OD trips were collected as well as travel time and signal settings. These data is in public as “Benchmark Dataset (BM)” [Hanabusa, et al. 2002] to be applied for the model validation.

So far, the validations of AVENUE [Horiguchi, et al. 1998] and NETSIM [Sawa and Yamamoto. 2002] with Kichijoji-Mitaka BM have been reported. Both of the cases compare the link throughputs from the simulation result with survey data and calculate the correlation coefficient ($R^2$) to evaluate the reproducibility of traffic condition. AVENUE was applied to the whole network that has alternative routes for each OD pair, then it gave quite satisfactory result as $R^2 = 0.98$. 
NETSIM, at first, was applied to the whole network, same as AVENUE. The reproducibility, however, was not satisfactory so that $R^2 = 0.67$, shown in the left plots of Figure 6. Furthermore, the linear regression line of the plots is slightly steeper than the diagonal line. This means NETSIM tends to overestimate the traffic volume when it is applied to the network containing loops [Sawa and Yamamoto, 2002].

Subsequently, NETSIM was applied to the corridor section in the network that has no alternative route for each OD pair. For this case, the reproducibility was improved ($R^2 = 0.90$) and the regression line also lies along the diagonal line.

The reason of this problem can be explained as follows. Since NETSIM does not incorporate drivers' route choice model, the traffic demand is given as turning volume ratio at each intersection. Thus, a vehicle may run along looped route and use the same link more than twice within its trip. This leads to overestimation of traffic volume. Therefore, the tester of NETSIM concludes that it should be applied only to corridor shape networks.

**FUTURE WORKS**

So far, we have been working for standardizing several stages in the utilization of simulation as well as the model development stage. These our activities can be found in “Clearing House” at an Internet website:
The developers and users of simulation models can also publish their experiences on verification and validation through Clearing House. Followings are the current menus available on Clearing House at present (some of them are still under construction, sorry). The best practice manual mentioned above will be included in the future.

- Introduction of traffic simulation models used in Japan.
- Verification results of the simulation models.
- Standard Benchmark Data Sets for Validation of Traffic Simulation Models.
- Validation result of the simulation models with BM data sets.
- Online Q&A.

We are now discussing how we encourage model developers to open their verification results to the public. Basically, we expect the verification process to be “de-fact standard” by educating the necessity of the verification to practitioners and also to people in public sectors who order consulting jobs using simulation models. The further discussion in our activity is expected that for to comprehend the results of the verification studies, and to estimate the characteristics of each model. Also, we will afford the movement of this standard certification process for other simulation models worldwide.

REFERENCES

'AVENUE Ver. 2', Proceedings of the Third Annual World Congress on Intelligent Transport Systems, Orlando.


