THE ART OF THE UTILIZATION OF TRAFFIC SIMULATION MODELS: HOW DO WE MAKE THEM BE RELIABLE TOOLS?

Ryota Horiguchi
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

Masao Kuwahara
Institute of Industrial Science, University of Tokyo
4-6-1-C504 Komaba Meguro-ku Tokyo 153-8505 Japan, phone: +81-3-5452-6419,
fax: +81-3-5452-6420, e-mail: kuwahara@nishi.iis.u-tokyo.ac.jp

ABSTRACT

This paper, at first, describes the current status of the utilization of traffic simulation models with the consideration based on a questionnaire survey for the model application in Japan. At second, the Best Practice Manual for Simulation Application, which is currently under preparation, is mentioned. Some exertions in the manual to address the issues in simulation application are presented in this paper; i.e. i) understand the models' nature through verification and validation, ii) OD estimation from vehicle counts, iii) model parameter calibration, and iv) indices to measure the reproducibility. At the end of this paper, Clearing House of Traffic Simulation Models that promotes the simulation utilization is introduced.

INTRODUCTION

This paper describes the art of the utilization of traffic simulation models in practical scene. In the last decade, quite a number of network traffic simulation models become available in all over the world. Many reports of the simulation studies can be found in business use now. 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 [Paramincs. 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 the 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.

Another criticism that "*simulation is useful, but dangerous* [Smartest. 1999]" could be pointed out. Case studies of traffic simulation are sometimes reported only with their results. However, there must be much room for the users to calibrate the simulation models out of given conditions. Since a simulation model has the flexibility to reproduce traffic conditions by changing its model parameters, users may fit the simulation result to "any" conditions. Different way of model calibration may lead to different result for the same case study, so that unless the report tells us how the model is fitted to the desired condition, we barely believe the result is pertinent.

In order to overcome such situation, Sim@JSTE is currently undertaking further exertions to establish the good manners for simulation studies as the "Best Practice Manual of Traffic Simulations". A simulation study consists of several phases to complete the application; e.g. data acquisition, model selection, model calibration, result interpretation, etc. We have been, at first, collecting the case studies for business practice through a questionnaire survey, and then extracting the generalized methodologies for each phase of application to be included in the manual.

In the following chapters, the current status of the utilization of traffic simulation models is, at first, described by examining the results of a questionnaire survey for the model application in Japan. At second, the best practice manual for simulation application, which is currently under preparation, is mentioned. Some exertions in the manual to address the issues in simulation application are presented in this paper; i.e. i) understand the models' nature through verification and validation, ii) OD estimation from vehicle counts, iii) model parameter calibration, and iv) indices to measure the reproducibility. At the end of this paper, Clearing House of Traffic Simulation Models that promotes the simulation utilization is introduced.

CURRENT STATUS OF SIMULATION UTILIZATION

Questionnaire Survey for the Simulation Application in Japan

In order to understand the realities of simulation model application, the questionnaire form shown in Figure 1 was delivered to the users who had much experience in simulation application [Horiguchi and Oneyama. 2002]. Similar questionnaire survey was achieved within Smartest Project in Europe [Smartest. 1997]. In the Smart Project survey, however, the major interest was to identify the requirements to the microscopic simulation models to be applied to the evaluation of advanced telematique services.

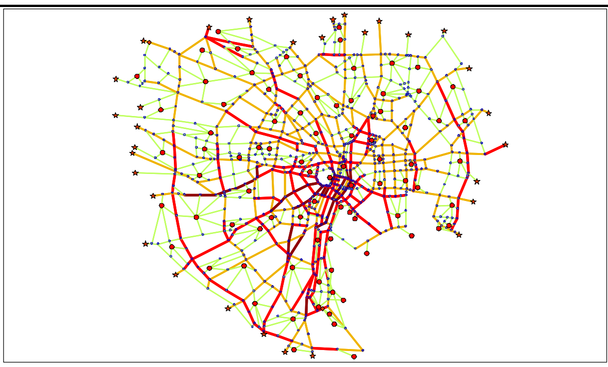
In order to understand some details of each simulation study, the following items were included in the questionnaire form:

i) *Outline of the case study* -- purpose of the simulation, evaluated measures, etc.
ii) *Network information* -- size, area, shape, road type, etc.
iii) *Input data & model parameters* -- trip demand, link capacity, scan interval, etc.
iv) *Reproducibility* -- how to calibrate the model, what's the indices, etc.
v) *Output measurement* -- sort of measurement, etc.

There are 41 forms concerning to 8 simulation models shown in Figure 2. AVENUE, tiss-NET, SOUND, VISITOK [VISITOK. WWW site], REST, and TRANDMEX [Suzuki, *et al*. 2000] are the models developed in Japan and the users are equal to the developers. Paramics and NETSIM are the overseas' models.

Figure 3 illustrates the rough classification of the models. The horizontal axis gives the spatial resolution of vehicle movement and the vertical axis shows the approximate size of applied network in usual. All models treat discrete image of a vehicle, i.e. no fluid approximation models, while there are two different types in the modelling of vehicle movement. "C-F type" represents the models based on "car-following behaviour", while "Q-K type" models use traffic flow characteristics such as "flow-density relationship" or "speed-density relationship". Although it is frequently mentioned that the "C-F type" is referred as the definition of "microscopic" model, we avoid using this word, because, as explained in the following section, it was difficult to find any significant differences in the usage of both "C-F type" and "Q-K type" for the "microscopic" case studies.

SIMULATION STUDY RECORD CARD

	Code	Traffic Management		**Application Type**	TDM	**Simulation Model**		SOUND	
Outline	**Subject**	Evaluation of Road Pricing Scheme in Tokyo Central Area							
Outline	**Purpose**	The purpose of this case is to evaluate the mitigation of traffic congestion with the road pricing scheme in Tokyo Central Area, which is currently investigated by Tokyo Metropolytan Government.							
Outline	**Note**	Estimate the future trip demand by considering human behaviour for the road pricing scheme.							
Network & Simulation Settings	**Area**	Tokyo central area (23 wards); Appx. 15km in radius				**Time Period**	24 hours		
Network & Simulation Settings	**When?**	Present - 1996, Future - not specified							
Network & Simulation Settings	**Type**	Major arterial roads							
Network & Simulation Settings	**Size**	# of Nodes	942	# of Links	2952	# of Centroids	115	# of Trips	約370万
Network & Simulation Settings	**Size**	Arterial Road Networks		# of Intersections	392	# of Traffic Signals	0*	# of Roads	708
Network & Simulation Settings	**Size**	Expressway Networks		# of JCTs		# of Ramps		# of Sections	
Network & Simulation Settings	**Note**								
Network & Simulation Settings	**Network Image**								

Input & Parameters	Road Network	Link	Length, number of lanes, capacity of lanes, free flow speed.
Input & Parameters	Road Network	Intersection	Capacity for each stream direction
Input & Parameters	Road Network	Junctions	
Input & Parameters	Signal	Cyc, Sp, Of	Signal control is not explicitly considered. Link capacities are reduced by the assumed signal split.
Input & Parameters	Signal	How do you know?	Based on assumption
Input & Parameters	Demand	Type	Time varying O-D matrix
Input & Parameters	Demand	How do you know?	Present: based on the result of Vehicle O-D Survey in the Traffic Census in 1996. Future: modified the presend OD with some human behaviour models.
Input & Parameters	Demand	Spacial Resolution	Middle Zone used in the Traffic Census.
Input & Parameters	Demand	Time Resolution	1 hour.
Input & Parameters	Demand	Vehicle Type	2 types (large / small)
Input & Parameters	**Note**		
Model Settings	**Scanning**		time periodic scanning (6 seconds for this case)
Model Settings	**Vehicle Size**		10 vehicles for each packet.
Model Settings	**Route Choice behaviour**		Stocastic route choice with Dial's algorithm. Travel time information is updated by 5 minutes. Pricing effect is incorporated into the generic cost.
Model Settings	**Note**		
Reproducibility	**Parameter Calibration**	What?	Capacity for each stream direction
Reproducibility	**Parameter Calibration**	How?	To fit the average link travel speed in Tokyo Central Area to the reference data.
Reproducibility	**Reference Data**		Average travel speed in Tokyo Central Area
Reproducibility	**Data Source**		Traffic Census in 1996
Output	In this case, the following output is evaluated; - average link travel speed, traffic throughput, areawise average travel speed.		

Figure 1: An Example of Model Application Form

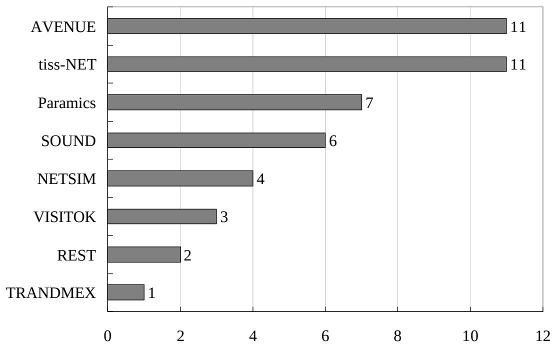


Figure 2: The Number of Collected Answers.

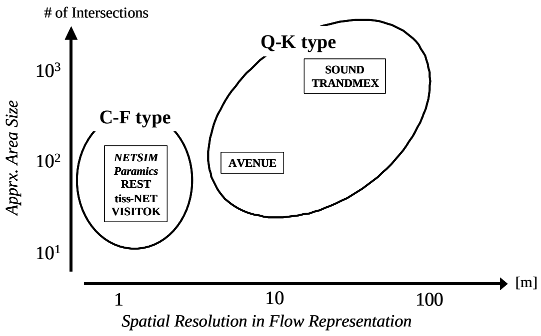


Figure 3: Rough Classification of the Models.

Outline of the Simulation Studies

Figure 4 illustrates the purposes of the simulation studies. Each model is denoted as the first two characters in its name. In order to distinguish the usage of "C-F type" and "Q-K type", the legends for the former are denoted as shaded columns, while the latter as hatched columns. Figure 4 tells us that the studies to be applied to smaller areas come to the fore, such as "Improve Bottlenecks", "Inter-modal Junctions", or "Building Shopping Malles". This may be come from not only that there are more so-called microscopic simulation models in the answers, but also that some technical difficulties in the studies with the larger areas, e.g. data acquisition, prevent from the applications.

Figure 5 sums up the countermeasures of case studies. The "C-F type" and the "Q-K type" are illustrated in the same way as Figure 4. The figure divides the countermeasures into "Operation and Management" and "Construction of Road and Parking Facilities". Similar to Figure 4, the local, i.e. microscopic, countermeasures are coming to the fore.

However it seems to be difficult to find the remarkable difference between those two types. Even if we use a simulation model of "Q-K type", we may apply it to some microscopic countermeasures as long as the model has sufficient temporal and spatial resolution for the concerning traffic phenomena. For instance, when we evaluate the effects of the improvement on signal control scheme, it may be enough when the model can reproduce the capacity as well as the shockwave propagation at every 1 or 2 seconds with the spatial resolution in several meters. Thus not only the "C-F type" models but also the "Q-K type" models can become "microscopic" models.

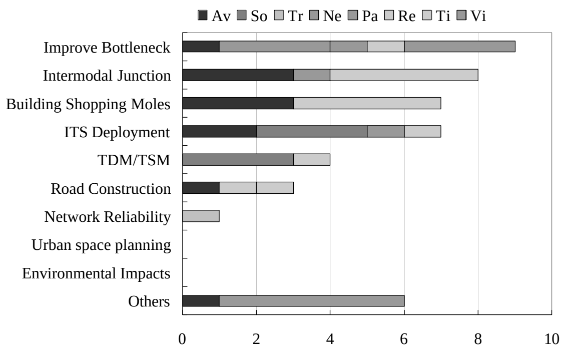


Figure 4: Purposes of the Case Studies.

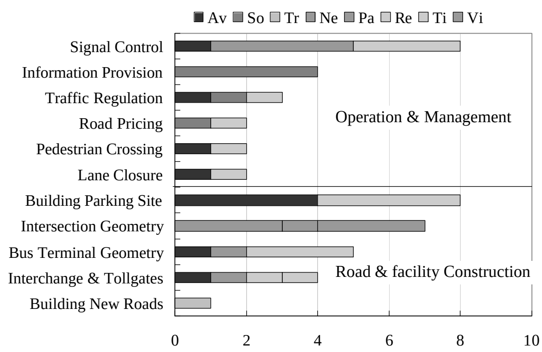


Figure 5: Countermeasures of the Case Studies.

Road Networks Applied in Simulation Studies

Figure 6 illustrates what road types are more commonly applied in the case studies. Most cases deal with the networks of surface streets so as to be applied with local countermeasures. This may come from the Japanese peculiarity that heavy congestions are frequently found on surface streets. However, we have to pay attention the fact there are no case studies with integrated networks of surface streets and expressways in practical business scene. There must be more or less requirements to the simulation to evaluate the macroscopic countermeasures like road pricing scheme, which we have to apply the simulation to considerably large areas including surface streets and expressways.

The biggest reason to this fact may come from that it is hard to know how the traffic demand arise and how the drivers choose the route on large networks. This sort of problems in data acquisition can be also implied from Figure 7; i.e. almost two third of the studies apply the simulation to the simple shape network that has no alternative route for each OD pair. The simulation users seem to be avoiding complicated network shape in practical scene. We will come back to the OD data acquisition problem in the subsequent section.

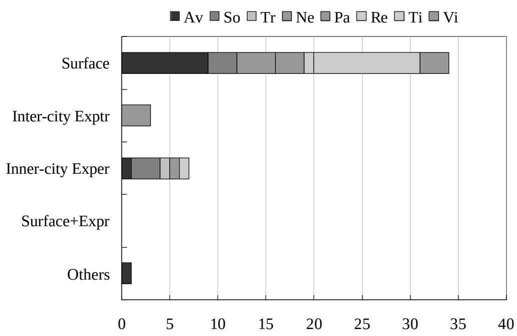


Figure 6: Road Types of the Simulation Studies.

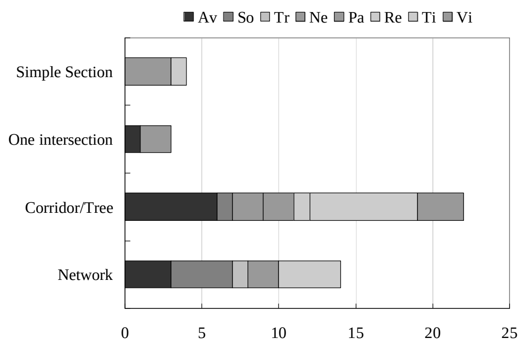


Figure 7: Network Shapes of the Simulation Studies.

Figure 8 compares the number of nodes to the size of each network measured by the approximate length of area boundary. It is natural that the number of nodes is getting larger as the size of network becomes large. In the mathematical expression of network, not only the subjected intersections or junctions but also dummy junctions can be thought as nodes.

Figure 9 now compares the number of actual intersections or junctions on the network. There are remarkable differences between the number of nodes and the number of actual intersections when the size of network is fairly large. One reason for this is considered that many dummy nodes are inserted into the network to connect with the zone centroids at which the traffic is generated and/or absorbed. Another reason is that many minor intersections are still remaining, of which either of crossing roads is not included in the subjective network.

Increasing the number of nodes (= the number of links) makes the simulation be difficult in the model calibration, since most of the models have the parameters to effect on the link capacity. It depends on the individual user’s skill that how to describe the network and how to calibrate the model. This may lead that the simulation result will be different for the same case with different users.

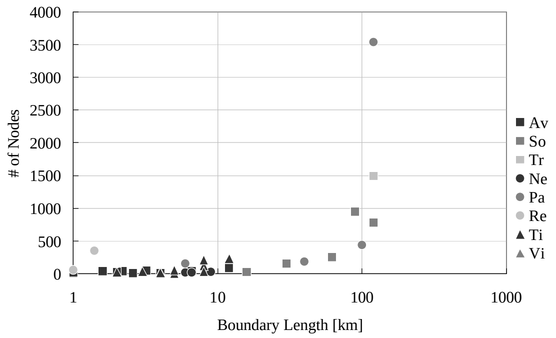


Figure 8: Network Size vs. Number of Nodes.

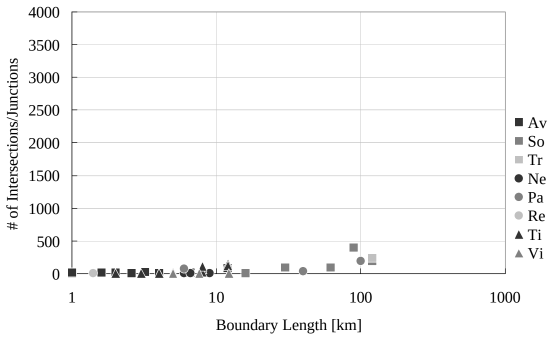


Figure 9: Network Size vs. Number of Intersections.

Traffic Demand Given to the Simulation Models

Figure 10 counts the data sources that the users used to obtain the traffic demand. Here two models (NETSIM and VISITOK) do not incorporate drivers' route choice model, so that they require the arrival flow rate at each network boundary and the turning ratio at an intersection.

The others accept time-varying OD matrices with incorporated route choice model, but the data sources are depend on the cases. Some of the cases with small networks use OD matrices directly observed by licence plate matching and some of them with large networks borrow the result of conventional survey such as traffic census. The former can obtain somehow precise OD matrices but it must be quite cost consuming way. The latter is cost effective but worse in preciseness because the sampling rate of the census is normally very small.

There are two items of "OD estimation" from vehicle counts, i.e. with the simple corridor networks and with complicated networks. For the corridor shape networks, we can calculate OD matrix by using the turning ratios at intersections. This is completely equivalent to the cases of "turning ratio" even if the simulation incorporates the route choice model.

OD estimation with complicated networks requires some technique. All cases here use the "extended entropy maximization method" [Oneyama, *et al*. 1996] to estimate time-varying OD matrix from traffic counts, but this sort of technique have not obtained popularity in practical scene yet. In the next chapter, we will survey some OD estimation technique.

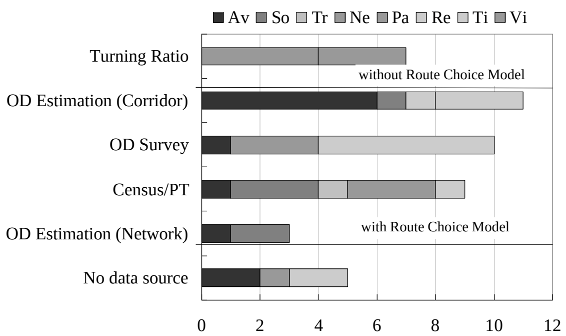


Figure 10: Data Sources of Traffic Demand.

Indices of Reproducibility of Traffic Conditions

Figure 11 shows what indices to be employed to evaluate the reproducibility of traffic conditions. In general, traffic condition on a link or a section must be explained with at least two measurements; i.e. throughput (TRPT) and travel time (TRVT), or throughput and queue length (QLEN). However, there are several cases using only one measurement because of the limitation of available data. For these cases, it is suspicious whether the simulation really reproduces the present condition of traffic.

For the cases with more than two measurements, queue length is more popular than travel time. The reason for this may be inferred from that it is hard to measure the travel time on overall network with sufficiently short interval, because the data collection mostly depends on small number of floating cars. On the other hand, the queue length involves some uncertainty because the definition of queue length seems to be unclear.

Further problem could be arise that how the users quantitatively estimate the reproducibility. A RMSE (root mean square error) or a correlation coefficient seems to be used from habit, but how do we decide the index value is sufficient or not? We will also consider this problem in the next chapter.

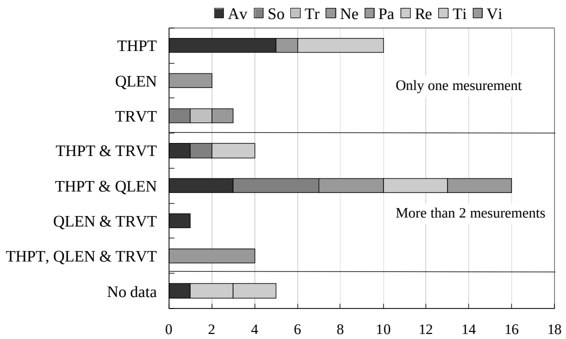


Figure 11: Indices of Reproducibility of Traffic Conditions

BEST PRACTICE MANUAL FOR TRAFFIC SIMULATION APPLICATIONS

Sim@JSTE is now working for the best practice manual for traffic simulation applications, which contains useful sample procedures for each issues in practical case studies. There are several issues in typical simulation applications; e.g.

i) Taking advantages of dynamic traffic simulation to the static analysis.
ii) Understanding models' nature through verification and validation.
iii) Traffic survey for dynamic traffic simulation.
iv) Trip demand estimation and network configuration.
v) Parameter calibration to reproduce present traffic conditions.
vi) Sensitivity analysis of each scenario.

For each issue, typical techniques that have been employed by the experts of simulation will be explained in the concrete. Not only the conventional techniques but also the advanced ones will be included in the manual. In this chapter, some of the contents are presented so as to grasp the outline of the manual.

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. In this section, let us introduce several results of verification and validation to understand how they help to comprehend the model characteristic [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 12 indicates 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, 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, and can give the meaningful implications that the 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.

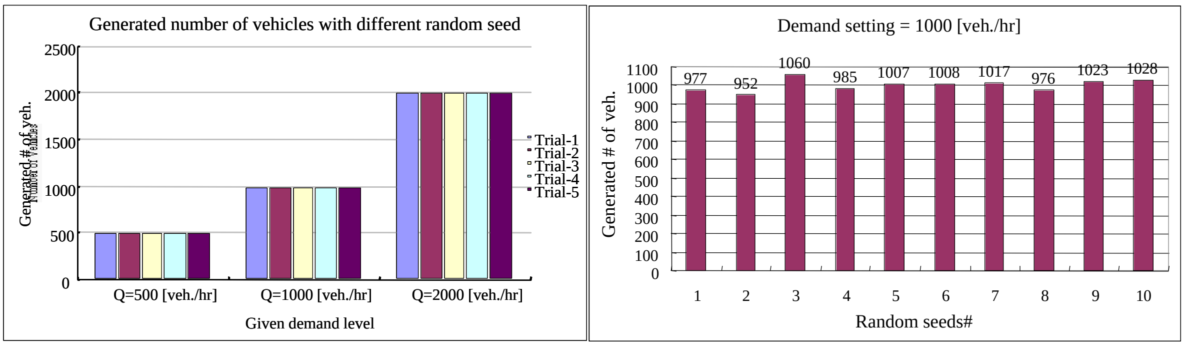


Figure 12: Total number of generated vehicles – AVENUE (left) and tiss-NET (right)

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.

Model	Driving behavior	Link performance
Paramics	minimum headway, maximum acceleration, driving aggression, etc.	headway coefficient, limit speed, gradient, etc.
SIPA	target headway, target speed, maximum acceleration, response delay, etc.	allowable minimum headway, limit speed, gradient, etc.

Table 1: Major model parameters of Paramics and SIPA

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 simulation result to an actual traffic condition by changing the most sensitive parameters.

Figure 13 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.

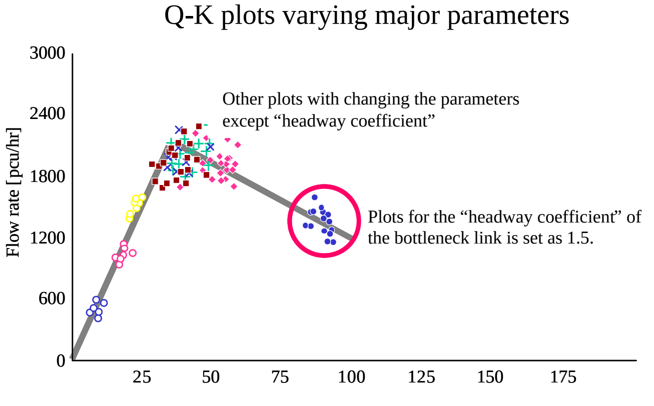


Figure 13: Bottleneck capacity and traffic flow characteristics of Paramics

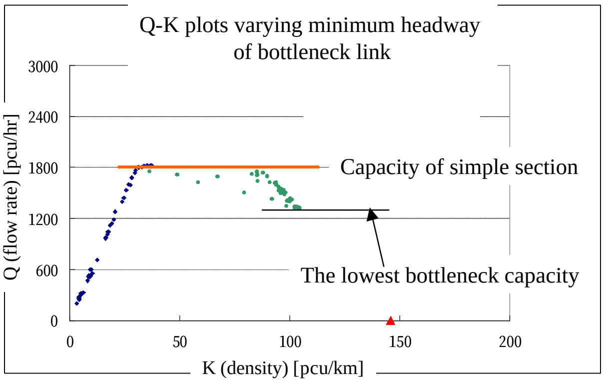


Figure 14: Bottleneck capacity and traffic flow characteristics of SIPA

Figure 14 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 C-F 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.

For the reason to iii), the tester of SIPA reports the study on the effects of the length of bottleneck sections. Figure 15 illustrates the reproduced bottleneck capacities that change as the length of the bottleneck link varies. The line with symbol indicates the achieved flow rate along the distance from the bottleneck section. In this case, the bottleneck link for each line has the same “minimum headway on the link” value, but has different length. The superimposed curve that lies on the flow rate at each downstream end of the bottleneck link shows that the capacity declines as the length of the bottleneck link gets longer.

This phenomenon may be considered as follows; when a vehicle with the minimum headway allowed in the upstream section enters the bottleneck link, the headway achieved by the vehicle does not immediately shift to the new minimum headway. If the length of a bottleneck link is so short that a vehicle passes through the link before its headway will shift to the new headway, the capacity of the bottleneck link will not decline as expected. The users of C-F type models should realize this phenomenon because it is found among most of the models, not only in SIPA.

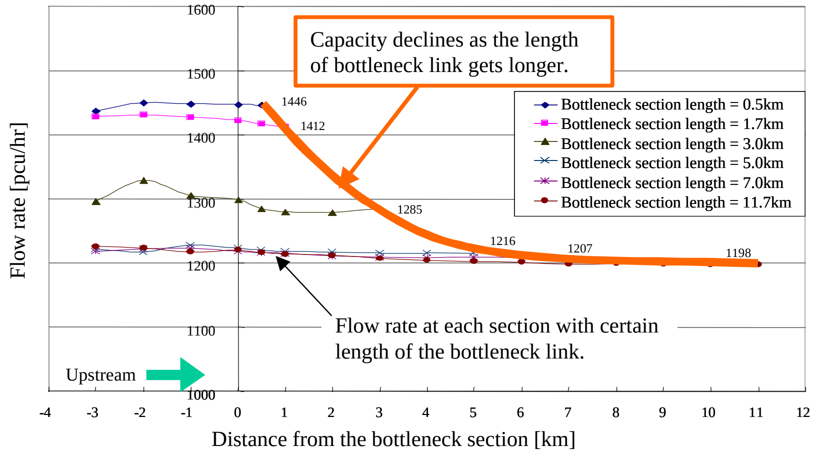


Figure 15: The capacity varying with the length of bottleneck section (SIPA)

<u>Saturation Flow Rate at a Signal Intersection</u>

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 16 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 (1600 pcu/G1hr) in average.

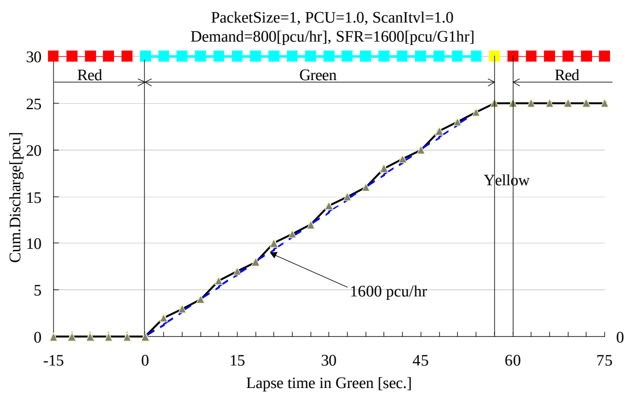


Figure 16: Link discharging profile at a signal intersection -- SOUND/A-21

There is another point to be discussed in Figure 16. 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 delay.

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 17 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 $R^2 = 0.98$.

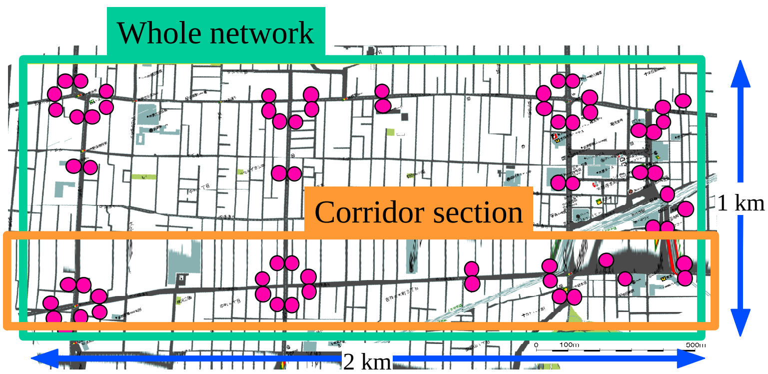


Figure 17: The Kichijoji-Mitaka Network Included in the Benchmark Dataset.

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 18. Furthermore, the linear regression line of the plots is slightly steeper than the diagonal line. This means NETSIM tends to over estimate 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 over estimation of traffic volume. Therefore, the tester of NETSIM concludes that it should be applied only to corridor shape networks.

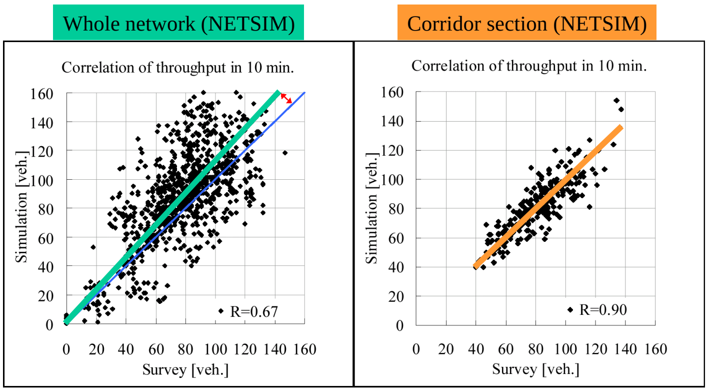


Figure 18: Link Throughput Reproducibility in the Validation of NETSIM [Sawa and Yamamoto. 2002]

Trip Demand Estimation

How do we obtain the trip demand to input traffic simulation can always be a big problem, especially for the model that accepts the trip demand in OD matrix. Some simulation models are combined with the software package to estimate OD trip demand, such as EMME/2 [EMME/2. WWW site]. However, most of those software basis on the estimation procedure used in conventional transportation planning; i.e. the four-step estimation method. Traffic simulation has less compliance to accept such OD trip demand, because it might be estimated without consideration of existing link flows at present. As long as using this sort of trip demand, it is difficult for the traffic simulation to reproduce present link throughput volume with satisfactory preciseness.

In order to reduce this impedance mismatch between the traffic simulation and the trip demand estimation, there are some research exertions that estimate OD trip demand using vehicle counts on links. Cremar and Keller [Cremar and Keller. 1987] propose the "entropy maximization" method that adjust the OD matrix so as to maximize the entropy calculated with given vehicle counts on link and assumed link choice probabilities for each OD pair. However, their method did not consider the dynamic aspects in traffic conditions, there still remains some mismatch to traffic simulation.

Oneyama, *et al.* [Oneyama, *et al.* 1996] extends the entropy maximization method to be applicable to dynamic traffic conditions. Their method re-constructs the road network representation along time-space axes as shown in Figure 19. According to the given traffic condition, the congested links are re-constructed as the three dimensional links with steep gradient in time-space. Observed vehicle count at each time slice on each link is assigned to each three dimensional link. By applying the entropy maximization method to this time-space network, we obtain the time-varying OD matrix with consideration of dynamic traffic conditions. Adding to this improvement, they incorporate the empirical OD pattern, which can be obtained by conventional questionnaire survey, to the calculation of entropy. This empirical OD pattern may affect the estimated OD matrix as restriction conditions. This "Extended Entropy Maximization Method ($(EM)^2$)" is now implemented in the software package to work with SOUND/A-21 [SOUND. WWW site].

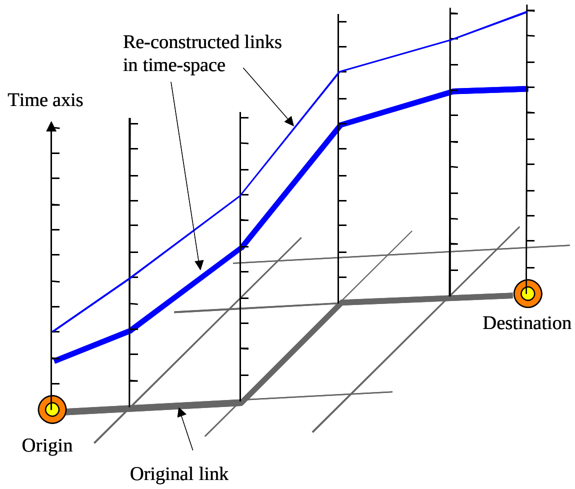


Figure 19: Re-constructed Time-Space Network Used for OD Estimation [Oneyama, *et al*. 1996]

Yoshii and Kuwahara [Yoshii and Kuwahara. 1998] work for the further improvement on $(EM)^2$. Since $(EM)^2$ calculates the choice probability of each three dimensional link by assuming certain stochastic route choice model, the reproduced link throughput in the simulation, which has different drivers' route choice model, may be different from the traffic counts used for the OD estimation. In other words, $(EM)^2$ adopts "predictive" route choice model while most of the simulation models use "reactive" route choice models.

The basic idea of their improvement is to replace 3-D "time-space" network with traffic simulation itself. By applying the simulation with some assumed OD matrix, it can provide the link choice probability that can be used for the calculation of entropy. On the contrary, the link choice probabilities may change when the assumed OD matrix is different. Therefore, the simulation and the OD estimation are iteratively executed and converge OD matrix so as to fit the link throughputs in the simulation to the observed vehicle counts. Kitaoka, *et al.* [Kitaoka, *et al*. 2002] employs the similar framework but decomposes time series in order to reduce calculation time and memory use.

Parameter Calibration to Reproduce Present Traffic Conditions

Most of the applications of traffic simulation include the phase to reproduce "present" traffic condition. The reason comes from not only the necessity for the basic case to be compared with the future cases, but also that we need the evidence that the simulation model we use has the ability to reproduce desirable traffic conditions by adjusting model parameters.

For the microscopic, i.e. small, case studies, it is enough to let the user calibrate the model parameters by hand, because the key parameters are not so many to confuse the user. On the other hand, for the macroscopic, i.e. large, case studies, there are too many model parameters to be adjusted by hand. We need some efficient methodology to tune these model parameters.

Cremar and Papageorgiou [Cremar and Papageorgiou. 1981] proposed the “Box Complex Technique”, which bases on random search technique to solve non-liner objective function. Nanthawichit and Nakatsuji [Nanthawichit and Nakatsuji. 2001] proposed the method to use Kalman Filtering that improve the estimation of model parameters by adjusting to the observed data. However, their methodologies are strongly conscious for the online application of traffic simulation and seem to be difficult to apply complicated networks except corridor shapes.

Furukawa, *et al.* [Furukawa, *et al.* 2000] proposed the method with heuristic rules to vary the link capacities of macroscopic simulation models. The reproduced condition in traffic simulation is transferred to the simple "point-queue" model. Each rule, which increases or decreases the link capacity, evaluates its agreement by operating the point-queue model. The rules that have the highest agreements compete other rules then change the link capacity in the simulation model. Repeating there processes converge the simulation result to the present traffic condition.

Indicators to Evaluate the Reproducibility

There is another problem in reproducing present traffic conditions, such that "how do we evaluate the reproducibility". We frequently use RMSE or correlation coefficient for the link throughputs and the travel times, but those indicators involve a couple of defects.

The first one is that they independently evaluate each time slice without considering time series. For the link throughput, the errors in earlier time slices may affect to the reproducibility of cumulative throughput much more than the errors in later time slices. The second is that their values include both the errors to the changes in long term and the ones of short fluctuations.

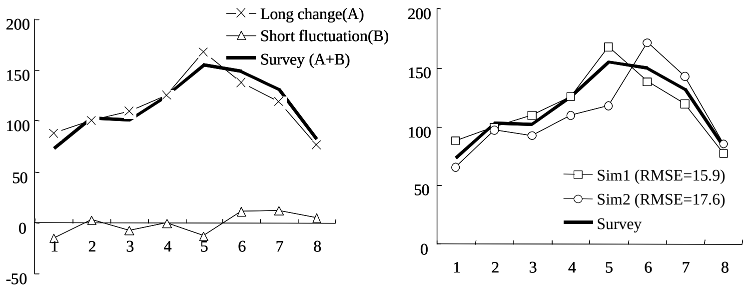


Figure 20: Problems for the Evaluation of Reproducibility with RMSE or Correlation Coefficient.

For instance, let us assume the survey values to be reproduced by the simulation, as shown in Figure 20, contain the long changes to be in subject and the short fluctuations not to be in subject. We have now two simulation results; Sim1 that completely agrees with the long changes in the survey data but does not agree with the short fluctuation, while Sim2 that does not agree with the long changes but completely agrees with the short fluctuation. RMSE value for the former is 15.9 and for the latter is 17.6, but how do we judge this tiny difference is significant or negligible? For this case, we finally have to approve the better one by "seeing" its shape on the graph. As subjective judgement lies here, the result of approval may depend on individualities and this situation is not desirable from the engineering point of view.

Horiguchi [Horiguchi. 2002] proposes the newer measurement index based on wavelet analysis, which is a popular technique to evaluate the "similarity" of two vectors. When the case to evaluate the similarity of the three simulated travel times to the observed value, as shown in Figure 21, the proposed method calculates the error component (= wavelet coefficient vector) at each frequency through the wavelet analysis. Figure 22 illustrates the basic idea of wavelet analysis. The original signal is decomposed into the approximation at lower frequency and the wavelet coefficient vector. Each wavelet coefficient vector represents the feature of changes in terms of frequency.

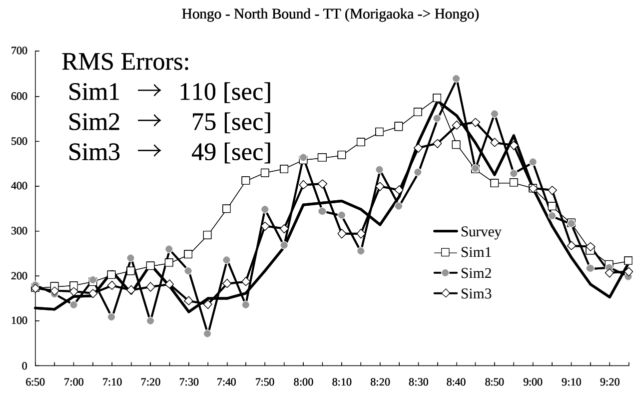


Figure 21: Examples of the Simulation Result.

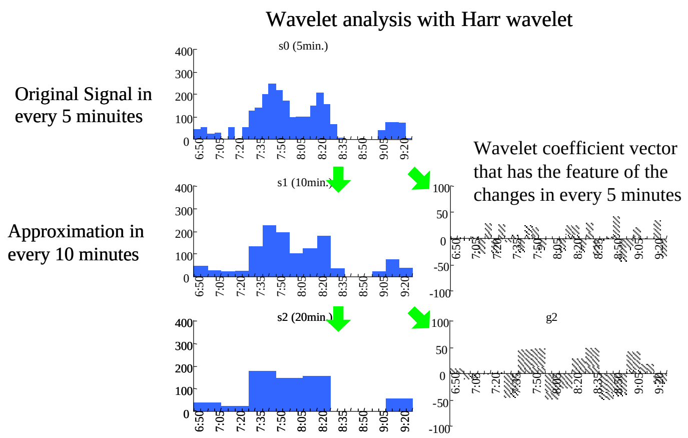


Figure 22: Basic Idea of Wavelet Analysis.

Figure 23 illustrates the RMSE of each wavelet coefficient vectors for each the simulated value to the one of observed value. As the RMSE at lower frequency (= longer cycle) gets small, two vectors can be thought to be similar in longer term. In general, the short term fluctuation can arise out of somehow microscopic traffic phenomena, which are often neglected in simulation. Therefore, we should select the allowable magnitude of error at each frequency in accordance with the simulation purpose, and normally, the allowable errors at higher frequency will be looser than the lower frequency.

For the example in Figure 22, Sim1 shows the larger errors than the others in lower frequency, i.e. Sim1 is no more similar to the observed value. Sim2 and Sim3 have the same errors except at the frequency in every 5 minutes. This means these two values well follow the observed value with the changes in longer terms, so that we may conclude Sim2 is satisfactory if we do not consider a great deal of every 5 minutes changes in the simulation.

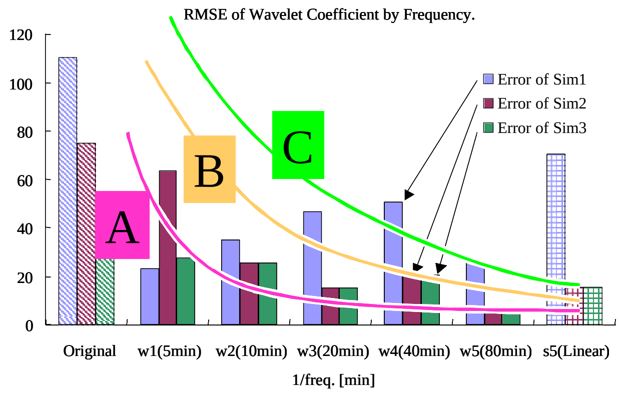


Figure 23: RMSE of Each Wavelet Coefficient Vectors at Each Frequency.

Our final goal is to put the curves, as displayed in Figure 23, to indicate the allowable errors with different grade. Here, the grade "A" means the strictest criterion while "C" means the looser one. What grade is to be satisfied may depend on the purpose of simulation application and the preciseness of available data. Some applications that require the reproducibility in short term changes, such as signal parameter optimisation, should satisfy the strict grade.

FUTURE WORKS THROUGH THE "CLEARING HOUSE" OF TRAFFIC SIMULATION MODELS

So far, we have been working for standardizing several stages not only in the model development but also in the utilization of simulation. These our activities can be found in “Clearing-House” at an Internet website:

http://www.jste.or.jp/sim/ (in Japanese, Figure 24).

The developers and users of simulation models can also publish their experiences on verification and validation through the clearing-house. Followings are the current menus available on the 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.
- Manual of Standard Verification Process for Traffic Simulation Models.
- 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.

Figure 24: The Clearing House of Traffic Simulation Models (in Japanese).

REFERENCES

AVENUE. *http://www.i-transportlab.jp/products/avenue/* (in Japanese).

Cremar, M. and Papageorgiou, M. (1981). Parameter identification for a traffic flow model, Automatica, **Vol. 17, No. 6**, pp. 837-843.

Cremar, M. and Keller, H. (1987). A new class of dynamic methods for the identification of origin-destination flows, *Transportation Research*, **Vol. 21B, No. 2**, pp. 117-132.

EMME/2. *http://www.inro.ca/products/e2fea.html*.

Furukawa, M. *et al*. (2000). Automatic tuning of parameters in a network traffic simulation model, *Proceedings of 7th World Congress on Intelligent Transport Systems*, Turin, CD-ROM.

Hanabusa, H. *et al*. (2001). Construction of a data set for validation of traffic simulations, *Journal of Japan Society of Civil Engineers*, **No. 688/IV-53**, pp. 115-123.

Horiguchi, R. *et al*. (1996). A network simulation model for impact studies of traffic management 'AVENUE Ver. 2', *Proceedings of the Third Annual World Congress on Intelligent Transport Systems*, Orlando.

Horiguchi, R. *et al*. (1998). A benchmark data set for validity evaluation of road network simulation models, *Proceedings of 5th World Congress on Intelligent Transport Systems*, Seoul.

Horiguchi, R. *et al*. (2000). A manual of verification process for road network simulation models – an examination in Japan, *Proceedings of 7th World Congress on Intelligent Transport Systems*, Turin, CD-ROM.

Horiguchi, R. (2002). What should be proper indicators to evaluate the reproducibility of traffic conditions for dynamic traffic Simulation? (in Japanese), *Proceedings of Infrastructure Planning*, June 2002, Nagoya.

Horiguchi, R. and Kuwahara, M. (2002). Verification process and its application to network traffic simulation models, will be published in *Journal of Advanced Transportation*, Special Issue of ITS.

Horiguchi, R. and Oneyama, H. (2002). Future direction for R&D activities of traffic simulation through the analysis on practical applications in Japan (in Japanese), will be published in *Journal of Japan Society of Civil Engineers*.

IATSS (2001). A manual of traffic assessment (in Japanese), *Report of International Association of Traffic and Safety Sciences*.

JSTE (2001a). A manual of standard verification process for traffic flow simulation model (Draft in English), *http://www.jste.or.jp/sim/verification/manual/VfyManE.html* (Japan Society of Traffic Engineers).

JSTE (2001b). Clearing house of traffic simulation models (in Japanese), *http://www.jste.or.jp/sim*.
Kitaoka, H. *et al*. (2002). Reproduction of traffic condition based on estimation of origin-destination flow (in Japanese), *Proceedings of Infrastructure Planning*, June 2002, Nagoya.
Nanthawichit, C. and Nakatsuji, T. (2001). Parameter estimation of macroscopic traffic simulation model, *Infrastructure Planning Review*, **Vol. 18, No. 4**, pp. 935-942.
NETSIM. *http://www.fhwa-tsis.com/*.
Oneyama, H., et al. (1996). Estimation of time dependent OD matrices from traffic counts, *Proceedings of the Third Annual World Congress on Intelligent Transport Systems*, Orlando.
Paramics. *http://www.paramics.com/*.
Sakamoto, K. *et al*. (1998). Traffic assignment method considering car-by-car behaviour for traffic impact studies - development of the tiss-NET system, *Proceedings of 8th World Conference on Transport Research*, Antwerpen.
Sawa, M. and Yamamoto, F. (2002). Validation of NETSIM using Kichijoji Benchmark Dataset (in Japanese), *Proceedings of Infrastructure Planning*, June 2002, Nagoya.
Smartest (1997). Review of micro-simulation models, *http://www.its.leeds.ac.uk/smartest/*.
Smartest (1999). Best practice manual, *http://www.its.leeds.ac.uk/smartest/*.
SOUND. *http://www.i-transportlab.jp/products/sound/* (in Japanese).
Suzuki, H. *et al*. (2000). A model development for estimations of origin-destination travel time and flow on freeways, *Proceedings of 6th International Conference on Application of Advanced Technologies in Transportation Engineering*, ASCE, CD-ROM.
VISITOK. *http://www.visitok.gr.jp* (in Japanese).
Yokochi, K. *et al*. (1999). Development of microscopic traffic simulator for AHS Evaluation, *Proceedings of 6th World Congress on Intelligent Transport Systems*, Toronto.
Yoshida, T. *et al*. (1999). A basic study for the planning of ETC-dedicated highway interchange using a traffic simulator REST (in Japanese), *Proceedings of Infrastructure Planning*, **No.22 (2)**, pp.231-234.
Yoshii, T. *et al*. (1995). An evaluation of effects of dynamic route guidance on an urban expressway network, *Proceedings of the 2nd World Congress on Intelligent Transport Systems*, Yokohama.
Yoshii, T. and Kuwahara, M. (1998). Estimation of a time dependent OD matrix from traffic counts using dynamic traffic simulation, *8th World Conference on Transportation Research*.