How Model Accuracy is Improved by Usage of Statistics?  
- An Example of International Freight Simulation Model in East Asia -

Ryuichi SHIBASAKI  
Dr. Eng., Senior Researcher  
Port and Harbor Department  
National Institute for Land and Infrastructure Management, MLIT, Japan  
3-1-1 Nagase, Yokosuka, Kanagawa  
239-0826, Japan  
Fax: +81-46-844-6029  
E-mail: shibasaki-r92y2@ysk.nilim.go.jp

Tomihiro WATANABE  
M. Eng., Head of Port System Division  
Port and Harbor Department  
National Institute for Land and Infrastructure Management, MLIT, Japan  
3-1-1 Nagase, Yokosuka, Kanagawa  
239-0826, Japan  
Fax: +81-46-844-6029  
E-mail: watanabe-t2w3@ysk.nilim.go.jp

Daishi ARAKI  
Researcher,  
Port and Harbor Department  
Pacific Consultants Co., LTD.  
Sekido 1-7-5, Tama City, Tokyo  
206-8550 Japan  
Fax: +81-42-372-6047  
E-mail: Daishi.Araki@ss.pacific.co.jp

Abstract: This paper shows how important equipment of statistics in Asian transport is for the model calculation in order to evaluate impacts of the policy to improve the environment of international freight transport. How the model accuracy is improved if some statistics are available in the model calculation, and how different the impacts to the model accuracy are according to the type of available data are discussed.

Key Words: international cargo flow simulation, East Asia, model accuracy, transport statistics

1. INTRODUCTION

Until now, the authors had developed a Model for International Cargo Simulation (MICS) which can simulate the movement of the cargo with the volume of OD container cargo as a given input, focused in East Asian region, in order to evaluate international freight transport policy. This model, for example, had already applied to simulate impacts on international cargo flow pattern in Eastern Asia as well as ASEAN countries, of investment for ASEAN Logistics Infrastructure Projects formulated by ASEAN Secretariat (2007a and 2007b) as prioritized projects.

However, as is well known, information and statistics on international transport and logistics in Asia are crucially lacking in every transport mode such as roads, railways, and maritime shipping. In order to calculate the model simulation, the authors have to estimate many input data from very scarce information by ourselves. In this paper, in order to show how important equipment of statistics in Asian transport is for this kind of model, how the model accuracy is changed if some statistics are not available in the model calculation is shown. Additionally, how different the impacts to the model accuracy are according to the type of available data is discussed.
2. ROUGH STRUCTURE OF THE MODEL

The detail of the structure of the model for international cargo simulation (MICS) developed by authors is explained in Shibasaki et al. (2009).

2.1. Model Overview

The original MICS focuses on the international maritime container cargo flow between all of the Japanese container ports, the major ports of East Asia, and other areas, as well as the land transport of container cargo within Japan, when the OD cargo flows (cargo demand) between regions are given. A structure of the model is shown in Figure 1. Shippers and carriers are assumed to independently optimize their different objectives or objective functions, and, as a result, reach to a Nash equilibrium under which they cannot improve them each other without changing a principle of the other party’s behavior.

![Figure 1: Structure of the MICS (Model for international cargo simulation)](image)

2.1.1 Shipper Sub-model

In the shipper sub-model, shippers maximize their utility individually, by selecting port for export and import, carriers for ocean-going shipping, and routes for land transport. The container cargo flow assignment in this sub-model is calculated, based on a nested logit type simulation; port and land route choice in the upper (first) level and carrier choice in the lower (second) level. In the first choice level, a Dial assignment, a typical stochastic network assignment methodology, is adopted instead of a plain logit model. In the second choice level for carrier selection, a simple logit model is applied and calculated logsum values in order to incorporate the world transport network for assignment in the first choice level.

i) upper level for export/import port and land route selection
\[ f_{rsk} = Q_{rs} \cdot \frac{\exp(-\theta \cdot c_{rsk})}{\sum_{k \in K_{rs}} \exp(-\theta \cdot c_{rsk})} \]  
\[ \text{s.t.} \quad c_{rsk} = \sum_{a \in A} \Lambda_{a} + \sum_{b \in K_{rs}} CL_{b} + \sum_{i \in K} CP_{i} \]  
\[ \Lambda_{a} = \frac{1}{\theta} \ln \left( \sum_{g \in G} \exp(-\theta \cdot CM_{ag}) \right) + \zeta \]  
\[ CL_{b} = CL_{cost_{b}} + vt_{shpr} \cdot CL_{time_{b}} \]

Here, \( f_{rsk} \): cargo volume of path \( k \) for OD-pair \( rs \), \( Q_{rs} \): OD cargo volume between \( r \) and \( s \), \( \theta \): variance parameter on Gumbel distribution, \( c_{rs,k} \): transport cost of path \( k \) for OD-pair \( rs \), \( K_{rs} \): choice set of path for OD-pair \( rs \), \( A_{a} \): expected minimum cost (logsum values) of international maritime link \( a \) included in path \( k \), \( CL_{b} \): generalized transport cost of land transport link \( b \) included in path \( k \), \( CT_{c} \): transshipment cost of inter-carrier transshipment link \( c \) included in path \( k \), \( CM_{agg} \): generalized cost of carrier group \( g \) in maritime link \( a \), \( G \): carrier group set (here, 9 groups), \( CL_{cost_{b}} \): monetary cost of link \( b \), \( vt_{shpr} \): value of time for shippers, \( CL_{time_{b}} \): transport time of link \( b \).

ii) lower level for carrier selection

\[ q_{ag} = d_{a} \cdot \frac{\exp(-\theta \cdot CM_{ag})}{\sum_{g \in G} \exp(-\theta \cdot CM_{ag})} \]  
\[ \text{s.t.} \quad d_{a} = \sum_{rs \in A} \sum_{k \in K_{rs}} \delta_{rs,k} \cdot f_{rsk} , \quad \forall a \in A \]  
\[ CM_{ag} = p_{ag} + vt_{shpr} \cdot CM_{time_{ag}} \]

Here, \( q_{ag} \): cargo volume of carrier group \( g \) in maritime route \( a \), \( d_{a} \): cargo demand of maritime route \( a \), \( \delta_{rs,k} \): Kronecker delta (:=1 when link \( a \) is included in path \( k \) for OD-pair \( rs \), :=0 otherwise), \( Q \): OD pair set, \( A \): maritime route set (because maritime route is defined by combination of a port for export and a port for import, the number of routes is equal to the number of ports multiplied by the number of ports minus one), \( p_{ag} \): freight for route \( a \) proposed by group \( g \), \( CM_{time_{ag}} \): total transport time for route \( a \) of group \( g \).

2.1.2 Carrier Sub-model

In the carrier sub-model, carriers maximize their net profits by making alliance of ocean-going shipping, by determining transport patterns of international maritime shipping including sizes of containership and choice of transshipment ports. Here, for the sake of simplicity, the algorithm of profit maximizing calculation is divided into two parts, i.e. revenue maximization and cost minimization, and these parts are independently calculated. The revenue of each alliance is maximized by routes consisting from one port for export to another port for import, by determining their freight of the route in question reflecting other carriers’ behavior and shippers’ behavior of carrier choice as described above (in the second level of shippers’ choice model). The transport cost of each alliance is minimized in total for the entire maritime shipping network; therefore, the system optimum (SO) assignment is adopted by carrier group.

i) profit maximization by maritime route

\[ \max_{p} \quad p_{ag} \cdot q_{ag}(p_{a1}, \ldots, p_{ag}, \ldots, p_{ag}) \quad \forall g \in G, \forall a \in A \]  
\[ \text{s.t.} \quad (5) \]
ii) cost minimization for entire maritime shipping network

\[
\min_x \sum_{v \in V} x_{vg} \cdot t_{vg}(x_{1g}, \ldots, x_{vg}, \ldots, x_{Gg}) \quad \forall g \in G
\]

(9)

Here, \( V \): link set of maritime shipping network, \( t_{vg} \): link cost function of link \( v \) of group \( g \). For the concrete definition of link cost function, please refer to other documents such as Shibasaki et al., 2005. For solving above maximization/minimization problems, because it is too difficult to solve explicitly, some kind of relaxation methods are adopted.

2.1.3 Convergence Calculation

A flowchart of the entire model calculation is shown in Figure 2. The model contains several convergence calculation steps. In addition to the two types of convergence calculation shown in the figure, both models of income maximization and total cost minimization of carriers also include convergence calculation respectively; therefore, four types of convergence calculation in total are hierarchically embedded in the model. An iterative calculation between 2) and 3) shown in the figure is aimed to obtain a Nash equilibrium in terms of carrier behaviors between income maximization including carrier choice behavior of shippers and total cost minimization, supposing a short-term optimization behavior. On the other hand, an iterative calculation between 1) and the above short-term optimization model of carriers (i.e. 2) and 3)) assumes a mid-term optimization behavior, obtaining a Nash equilibrium between port choice behavior of shippers and net profit maximization behavior of each carrier.

![Figure 2 Calculation flowchart of the MICS](image-url)

2.2 Data Preparation

The input data required for the model can be divided into four types; 1) amount of OD container cargo by region; 2) transport network data such as physical distance and operational costs including both international maritime shipping and hinterland transport; 3) service level of each port, such as the number of berths by depth, various fare and costs associated with berthing and sizes; and 4) initial input such as link flow between ports by ship size and by carrier and total volume of containers handled by port. This model was basically developed based on year-2003 data, where the latest data are available such as Survey Report of
International Container Cargo Flow, conducted by Japanese government every five year. These four types of data are all difficult to be obtained, especially the amount of container OD data.

The first model was developed under a network including 17 Japanese container ports, 8 Chinese ports, and 25 other Asian ports. The number of category in containership size was four (under 1000 TEU, 1000-2500 TEU, 2500-4000 TEU, over 4000 TEU). In addition, Japanese land transport network was only included as hinterland transport network; therefore, import/export port choice behavior of shippers was only considered for Japanese cargo.

2.3 Model Solution and Accuracy

Out of many exogenous variables in the formulations of the model, variance parameter on Gumbel distribution included in a logit model, $\theta$, and money-cost conversion factor (value of time for carriers) in the carrier sub-model, $v_{t,carr}$, are the most difficult to be exogenously given for optimal values among others, because data accumulation on the actual values of them are inadequate at all. Therefore, the optimal values of these parameters are decided to reproduce the actual container cargo flow pattern in the model, by applying a steepest descent method which is one of optimization methodology.

The reproducibility of the container cargo throughput for export/import is shown in the left side of Figure 3. Here, since shippers’ port choice behavior is only considered for Japanese cargo in the first model, the comparison of the actual and estimated volume is shown only for Japanese ports. As shown in the figure, container cargo throughput for export/import was reproduced well by the developed model in the most of Japanese ports. The right figure of Figure 9 shows the reproducibility of the volume of transshipped container cargo in all Asian ports. As shown in the figure, they were also reproduced well in the most of Asian ports.
2.4 Model Extension to Eastern Asia

2.4.1 Outline
In order to do simulations in the ASEAN Countries, hinterland transport network in East Asia needs to be incorporated in the above developed model as well as in Japan. The network incorporated includes road, railways, and ferry as shown in Figure 4. The total length of the network incorporated more than 333,000 km network with 4,885 links even in South East Asia. Some container ports and ferry ports were also added to be 33 container ports and 11 additional ferry ports in the entire ASEAN countries.

2.4.2 Data Preparation
For making OD cargo matrix in regional basis, cargo demand in country basis estimated by the trade model needed to be proportionally divided in principle according to the share of magnitude of regional economy such as GRP (gross regional products) by statistics issued by each government (e.g. Department of Statistics Malaysia, 2001, Statistical Yearbook Thailand 2004, Socio-Economic Statistical Data of 64 Provinces and Cities by Statistical Publishing House Ha Noi, 2005, and Statistical Yearbook of Indonesia, 2004).

2.4.3 Model Accuracy of the extended model
The reproducibility of the extended model in terms of the container cargo throughput for export/import in all Asian ports is shown in the left side of Figure 5. As shown in the figure, container cargo throughput for export/import was also reproduced well by the developed model in the most of Asian ports, except for some ports such as port of Guangzhou and Shenzhen in China. The right figure of Figure 5 shows the reproducibility in terms of the export/import container cargo throughput in only ASEAN ports. Also, estimated international cargo flow on the land transport network is shown in Figure 6.
Figure 5 Reproducibility of the extended model in terms of export/import (left) and transshipped (right) container cargo throughput (in 2003)

Figure 6 Estimated international cargo flow on the land transport network (in 2003)
3. DIFFERENCE IN SIMULATION RESULTS BY DATA AVAILABILITY

In order to examine the effects of statistics, the authors prepared three trial calculations to compare the results with and without considering some specific data, i.e., Chinese Custom Statistics (CCS) Data, regional economic data in Laos and Cambodia, and regional population data in Myanmar.

3.1 Difference by Incorporating China Custom Statistics Data

In CCS Data, trade values of Chinese cargo by location of customs (nearly equal to by Provinces) by partner countries are available by commodities (HS 2 digit) and by transport mode. In the case utilizing CCS data (with CCS case), OD cargo matrix predicted based on extended GTAP model developed by Shibasaki et al. (2008) by seven regions of China are divided into province basis (31 zones) with keeping the shares of provinces by partner countries. In the case not utilizing CCS data (without CCS case), OD cargo matrix predicted by seven regions are used without modification.

The simulation result of both cases as well as the actual in export/import container cargo throughput in Chinese ports is shown in figure 7. In most ports such as Tianjin, Qingdao, Ningbo, Fuzhou, Xiamen, PRD ports, and Hong Kong, estimated throughputs in with CCS case are closer to the actual throughput than those in without CCS case. In addition, Table 1 shows the predicted number of trucks in 2003 crossing national land border from/to China. Although the actual data crossing national border is not available, predicted flows to/from Hong Kong and Vietnam from/to China are increasing by incorporating CCS data. According to interview survey and fragmentary filed trip by the authors, these estimated results are possibly still underestimating the actual, thus it might be closer to the actual by incorporating CCS data.

![Figure 7 Predicted export/import container cargo throughput in Chinese ports (in 2003)](image)
Table 1 Predicted international cargo flow (number of trucks in 2003)
crossing national land border from/to China

<table>
<thead>
<tr>
<th></th>
<th>with CCS</th>
<th>without CCS</th>
<th>increasing ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong</td>
<td>1,395,147</td>
<td>1,157,839</td>
<td>20.5%</td>
</tr>
<tr>
<td>N. Korea</td>
<td>89,967</td>
<td>89,961</td>
<td>0.0%</td>
</tr>
<tr>
<td>Mongolia</td>
<td>54,975</td>
<td>54,975</td>
<td>0.0%</td>
</tr>
<tr>
<td>Vietnam</td>
<td>361,742</td>
<td>225,990</td>
<td>60.1%</td>
</tr>
<tr>
<td>Lao</td>
<td>9,899</td>
<td>9,870</td>
<td>0.3%</td>
</tr>
<tr>
<td>Myanmar</td>
<td>1,135</td>
<td>1,129</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

3.2 Difference by Incorporating Economic Data in Laos and Cambodia

Normally, in Laos and Cambodia, there are no available statistical data on regional economy such as GRP and trade value by provinces. The only data available in regional basis are demography in both countries; however, it is broadly known that geographical distribution of population is very different from a level of regional economy, especially in developing countries. Fortunately, the authors could access another regional economic data; GRP (in 2002) in Laos estimated by JICA on trial basis, and the number of garment factory by regions in Cambodia, which is considered more representative for the regional economic power. Table 2 shows the regional import/export cargo volume by aggregating model input (OD matrix) for each estimation method that divides bilateral trade amount into regional shares of each index. From the table, the regional cargo demand estimated by economic indices such as GRP and the number of factory has a tendency to concentrate into central region of the country, compared with those estimated based on the demographic statistics.

Because the amounts of cargo in these countries are small, a trend in the world international cargo flow such as cargo throughput of each port is scarcely changed. However, land flows of international cargo around these countries are slightly different especially in Cambodia as shown in Figure 8. For example, because there are very few industries in the northwest area of Cambodia despite much population, international cargo flow calculated based on the regional population could be overestimated. In fact, cargo flow crossing national border with Thailand calculated based on the economic indices is estimated almost half (10,016 unit/year) than that calculated based on the demographic statistics (19,326). Instead, the cargo flow between port of Sihanoukville, the biggest ocean container port in Cambodia, and Phnom Penh, the capital, and that crossing national border with Vietnam calculated based on the economic indices (130,335 and 13,191, respectively) are larger than those calculated based on the demographic statistics (114,452 and 8,921).

Table 2 Regional import/export cargo volume in Laos and Cambodia for each estimation method (TEU, in 2003)

<table>
<thead>
<tr>
<th></th>
<th>Estimated by Economic Indices</th>
<th>Estimated by Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Export</td>
<td>Import</td>
</tr>
<tr>
<td>Laos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>6,873</td>
<td>17,422</td>
</tr>
<tr>
<td>Central</td>
<td>12,065</td>
<td>30,583</td>
</tr>
<tr>
<td>South</td>
<td>3,872</td>
<td>9,816</td>
</tr>
<tr>
<td>Cambodia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North East</td>
<td>1,145</td>
<td>2,585</td>
</tr>
<tr>
<td>North West</td>
<td>286</td>
<td>646</td>
</tr>
<tr>
<td>Central</td>
<td>75,553</td>
<td>170,623</td>
</tr>
<tr>
<td>South East</td>
<td>286</td>
<td>646</td>
</tr>
<tr>
<td>South West</td>
<td>3,148</td>
<td>7,109</td>
</tr>
</tbody>
</table>
3.3 Difference by Regional Division of Myanmar by Population

In Myanmar, as well as situation in Laos and Cambodia, there is no economic statistics at all on a regional basis. Therefore, it was not divided into a regional level in the model developed so far. In this section, the model calculation with dividing Myanmar into 14 states and divisions, by the demographic statistics on a regional basis, although it is very rough approximation of the regional economic power as discussed in the previous section.

As well as the previous section focusing on Laos and Cambodia, the division of Myanmar does not also bring any significant change in terms of cargo throughput in each port. However, as will be understood, as shown in Figure 9, the estimated land flow of international cargo in Myanmar become more detail and spread over the country, compared with that calculated without regional division shown in Figure 6. In addition, the estimated cargo flows crossing national border of Myanmar are also increasing by the regional division; for example, the cargo flows across the border with China are estimated totally 12,038 unit/year, while they are estimated only 1,135 before division.
4. DISCUSSION

Until the previous chapters, the authors examined the difference of simulation results in international cargo flow by incorporating different or additional data sources, especially focusing on the accuracy of the cargo demand (OD cargo matrix) as model input using various economic data. It is no doubt there are some differences in the results, but how important these differences are? When discussing the accuracy of the cargo demand according to the results in the previous chapter, there could be summarized into two major viewpoints; detail zoning and usage of more realistic data.

*Detail zoning* seems to be very useful to enhance the model accuracy, as shown in Chinese case described in 3.1 and Myanmar case in 3.3. Although depending on the degree of detailing, in some case it could improve the accuracy of the cargo throughput handled in ports like Chinese case, not only the land flow of international cargo. In addition, the results of detailing Myanmar described in 3.3 implies that detail zoning are very significant influences on the estimation results of international cargo flow even if the cargo demand are divided according to the geographic distribution of population. That is to say, it means “better than nothing”.

On the other hand, the impact of *usage of more realistic data* seems to be limited, especially in terms of on worldwide international cargo flow, according to the results described in 3.2. It
could attribute to the fact that the cargo demand in Laos and Cambodia are very small in the world, but partly due to that in our model these economic indices are utilized only in order to calculate the share of each region. However, its impact on the land flow of international cargo crossing national border is significant, as well as detail zoning, because the cargo flow in each border tends to depend on the demand of the nearest zones.

Additionally, both detail zoning and using more realistic data bring significant difference when evaluating policy effects such as infrastructure investment and alleviating barriers of national borders. For example, Myanmar government put priority to develop Kyaukpyu Deep Sea Port located in the northwest coast (which is one of additional projects of ASEAN Logistics Infrastructure Projects formulated by ASEAN Secretariat, 2007); however, in our model before the regional division, the effect of the development cannot be measured precisely because all of Myanmar cargo assumed to be originated from/destined into Yangon, which is the only one origin/destination node in Myanmar, and most of them are imported into/exported from Port of Thilawa, situated closer to Yangon. According to the trial calculation using our model, when considering the regional division in Myanmar, predicted container cargo throughput in year 2020 of port of Thilawa and Kyaukpyu are 610,064 and 122,348 TEU respectively, while they estimated 704,968 and 58,081 TEU without regional division.

Different data source also make different results in terms of the benefit of the project. For example, the decreased amount of the transport cost of international cargo from/to Cambodia by implementing all of ASEAN Logistics Infrastructure Projects located in Cambodia is estimated 695 million US dollar per year according to the calculation based on the regional economic data of the country, while that according to the calculation based on the regional demographic data is estimated 1,271 million US$. This big difference may attribute to the fact that the effect of the targeted projects is larger in the north area of the country, and that how much cargo demand is estimated in these area are very different among estimation methods.

In this paper, the authors only focus on the accuracy of cargo demand in order to improve the accuracy and stringency of the developed model. Because cargo demand is the most important input of the model, the authors must keep improving its accuracy, and at the same time, detailing their zones as much as possible. Simultaneously, various parameters in the cost function and data on transport network and infrastructure including maritime shipping, roads, and railways also need to be validated and improved at any time.

ACKNOWLEDGEMENTS

This research was partly supported by Grant-in-Aid for Scientific Research (C) No. 21560565.

REFERENCES

American Digital Cartography incorporation (2005), ADC WorldMapTM Version 4.0
ASEAN Secretariat (2007a), Policy and Development Framework Report for ASEAN Logistics development Study
ASEAN Secretariat (2007b), Fact and Assessment Report for ASEAN Logistics development Study
Department of Statistics Malaysia (2001)
Global Insight Inc., World Trade Service Data
Goodwill China Business Information Limited, China Customs Statistics 2003
Statistical Publishing House Ha Noi (2005), Socio-Economic Statistical Data of 64 Provinces and Cities