Trip Distribution Modeling in Tokyo Bay Based on AIS Data

El-Hocine TASSEDA 1 and Ruri SHOJI 2

Abstract
In traffic modeling, a model is often assembled to simulate vessel streams within a designated traffic analysis zone, hereafter referred to as TAZ, such as harbors and straits. The traffic model usually consists of sub-models concerned respectively with traffic generation, and trip distribution.

In this paper, the underlying behavior and distribution of ships navigating from an Origin TAZ to a Destination TAZ are analyzed to assess the parameters behind the destination attribution based on the amount of the generated traffic at the origin. Then, the uncertainty associated and randomness with the traffic movements is estimated based on the concept of Entropy.

The results of this work provides an effective tool for evaluating the distribution of vessel traffic streams loads and appraising the level of disorder caused in Tokyo Bay. Furthermore, the model formalizes the trip distribution into a matrix that can be used as a metric for traffic generation and evaluating the fluctuation in traffic and/or TAZ traffic load assignment. Nevertheless, further research is needed to assess the traffic within the same TAZ by breaking-down every TAZ into Sub-TAZs.

Keywords: traffic modeling, trip distribution, traffic distribution, entropy

1. Introduction
In traffic modeling (1)(2)(3)(4)(5), a model is often assembled to simulate vessel streams within a designated traffic analysis zone, hereafter referred to as TAZ, such as harbors and straits. The traffic model usually consists of sub-models concerned respectively with traffic generation, and trip distribution (6)(7).

Trip distribution analysis is the process by which ships sailing from one TAZ to another TAZ are analyzed, and the trip distribution model is used to distribute the generated traffic from an Origin TAZ to a Destination TAZ. Compared to land traffic trip distribution analysis --where many models are developed based on traffic purpose, route, and costs related to every destination-- maritime traffic is constrained by the choice of the route, the destination based on vessel type and cargo type, and above all, the capacity of the TAZ.

In this work, the underlying behavior and distribution of ships navigating from an Origin TAZ to a Destination TAZ are analyzed to assess the parameters behind the destination attribution based on the amount of the generated traffic at the origin. Then, the uncertainty associated with the traffic movements within the designate TAZ is estimated based on the concept of Entropy (7)(8)(9)(10).

2. Analysis data
Tokyo Bay historical AIS (11)(12) data is collected from the Tokyo University of Marine Science and Technology Advanced Navigation System as follows:

Data Type: AIS data.

---

1 正会員 NYK LNG シップマネジメント株式会社 (〒100-0005 東京都千代田区丸の内 2-3-2 郵船ビル)
tassedahocine@gmail.com

2 正会員 東京海洋大学工学部 (135-8533 東京都江東区越中島 2-1-6)
- Time of Analysis: From 00:00 to 24:00.
- Type of Ships: All types.
- Number of Plotted Ships: 4521 plots (All Tokyo Bay)
- Analyzed Trips: 3410 trips (Trips within the defined TAZs).

3. Analysis area

Tokyo Bay, with its 1320 km$^2$, is selected as an analysis area. There are eight port areas in Tokyo Bay namely: Yokosuka Port, Yokohama Port, Kawasaki Port, Tokyo Port, Funabashi Port, Chiba Port, Anegasaki Port, and Kisarazu Port as shown in Fig.1(a). Based on the port limits introduced above, nine TAZs are determined for this analysis as shown in Fig.1(b). The defined TAZs areas are as follows:

1. Tokyo Bay Line TAZ: between the Tokyo Bay Line and Line “A”;
2. Yokosuka TAZ: Northern area of Yokosuka Port. The southern part is excluded from the analysis because of the ferry traffic and traffic around the Uraga Pilot Station which do not have any relevance to this work.
3. Yokohama TAZ: delimited by Yokohama Port;
4. Kawasaki TAZ: delimited by Kawasaki Port;
5. Tokyo TAZ: delimited by Tokyo Port;
6. Funabashi TAZ: delimited by Funabashi Port;
7. Chiba Port TAZ: delimited by Chiba Port;
8. Anegasaki TAZ: delimited by Anegasaki Port;

4. Research method

AIS destination data is one of the dynamic data that every ship should update after departure from port, and when destination changes. As per the Japanese Port Regulation, all ships --sailing into a port or in the vicinity of its boundary for the purpose of entering the port to which the Act on Port Regulation applies-- should enter the destination code designating the destination port in the column for information on destination as summarized in Table 1.

Following the above, port destination data is extracted from the AIS data. Unfortunately, the data is not accurate and the rate of non-compliance with the entry method is high. In addition, AIS data does not include the origin port necessary for our analysis. Due to this, it has been concluded that the AIS data is not suitable for extracting destinations and origins.

4.1 Problem statement

The problem --in this part of the work-- is to track and count all ships navigating from an Origin TAZ to a Destination TAZ.

4.2 Tracking algorithm

To solve the destination and origin issues described previously, an algorithm is developed to extract the Origin TAZ and Destination TAZ data from ships’ tracks. The algorithm is built on the assumption that every ship sailing in Tokyo Bay has a start position
where the ship appears on data screen, and an end position where the ship lays still in the water after berthing or anchoring. The algorithm tracks every ship sailing from one TAZ to another, and excludes any other ship that is not provided with an Origin TAZ and/or Destination TAZ.

4.3 Algorithm test

Fig. 2 shows the result of some trip analysis carried out for the AIS data on the 11th of November 2011. The light gray (O) pushpins indicated the Origin TAZ, while the dark gray (D) pushpins indicate the Destination TAZ.

where:
(a): Traffic from Tokyo Bay Line TAZ to all other TAZs.
(b): Traffic from all TAZs to Tokyo Bay Line TAZ.
(c): Traffic from Tokyo Bay Line TAZ to three selected TAZs, namely: Tokyo TAZ, Yokohama TAZ, and Kisarazu TAZ.

The test results show the adequacy of the built tracking algorithm for the purpose of this analysis.

5. Trip distribution model

Assuming \(S\) is the number of ships calling the analyzed area and \(N\) is the number of TAZs, ships departing from \(TAZ_i\) are indexed as \(A_i\) with a probability \(u_i\), and ships arriving at \(TAZ_j\) are indexed as \(B_j\) with a probability \(v_j\). Ships sailing from \(TAZ_i\) to \(TAZ_j\) are represented by \(f_{ij}\) with a trip probability \(\rho_{ij}\).

\[
\begin{align*}
\Sigma f_{ij} &= A_i \\
\Sigma B_j &= \Sigma A_i = S \\
\rho_{ij} &= \frac{f_{ij}}{S}
\end{align*}
\]

As for the probability of occurrence, it is defined as:

\[
\begin{align*}
u_i &= \frac{A_i}{S} \\
v_j &= \frac{B_j}{S}
\end{align*}
\]

where:

\[
\begin{align*}
\Sigma u_i &= \Sigma v_i = \Sigma \rho_{ij} = 1 \\
0 &\leq \rho_{ij} \leq 1
\end{align*}
\]

Table 2 illustrates the trip distribution model matrix as defined previously.

The uncertainty on the destination and origin can be reduced provided that the constraints imposed on every TAZ are known beforehand, such as the type of ships allowed, the maximum allowable dimensions (Length Over All, Breadth, and Draft), allowable type of cargo, and available handling infrastructure. Unfortunately such information is not available, and so the uncertainty associated with the traffic movements is estimated based on the concept of Entropy.

The Entropy \(H\) is defined as the uncertainty associated with the traffic distribution within a specific area, and it is related to the probability distribution of generated trips between the origin and destination. For the above mentioned trip distribution model, the Entropy \(H\) is defined as:

\[
H = -\Sigma_{ij} \rho_{ij} \ln(\rho_{ij})
\]

where:
\[
\begin{align*}
\rho_{ij} = 0 & \Rightarrow H = 0 \\
\rho_{ij} = 1 & \Rightarrow H = 0 \\
0 \leq \rho_{ij} \leq 1 & \Rightarrow H > 0
\end{align*}
\] (7)

Entropy H is generally thought of as a metric of a system’s state of disorder, where the higher a system’s entropy is, the more the system is disordered. And generally systems tend to move toward higher entropy values-- at which system stabilization is sought.

Similarly to the trip distribution model matrix, an entropy distribution matrix can be generated based on Equation 6 and Equation 7. The Entropy distribution model, unlike the trip distribution model described in Table 2, is not a behavioral model. That is, the Entropy distribution model does not strive to predict the trip distribution by modeling the ships’ behavioral aspects related to choosing a destination. The model on the other hand, attempts to determine a distribution of trips which are most likely to occur assuming that each trip occurs independently of the others; allowing us to assess the uncertainty behind destination attribution based on the generated traffic at the Origin TAZ.

6. Results

Table 3 shows the trip probability distribution of all trips made from TAZ\textsubscript{i} to TAZ\textsubscript{j} for the time domain covering the period of analysis and where no constraint is imposed. From Table 3, it can be seen that the traffic originating from TAZ\textsubscript{i} towards a destination TAZ\textsubscript{j} where \( i = j \) (underlined values) accounts for a trip probability \( \rho = 0.34868 \).

Fig. 3 shows two similar cases for Yokohama TAZ and Tokyo TAZ. The traffic between same TAZs is attributable to service boats, leisure passenger ships and the like.

Vessel traffic streams navigating in the same TAZ are not relevant for our analysis since it can be only quantified but not reconstructed through modeling.

Table 3 Tokyo Bay trip distribution model without constraint

<table>
<thead>
<tr>
<th>TAZ\textsubscript{j}</th>
<th>Tokyo Bay Line</th>
<th>Yokosuka</th>
<th>Yokohama</th>
<th>Kawasaki</th>
<th>Tokyo</th>
<th>Funabashi</th>
<th>Chiba</th>
<th>Anegasaki</th>
<th>Kisarazu</th>
<th>( \sum TAZ\textsubscript{i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo Bay Line</td>
<td><strong>0.0648</strong></td>
<td>0.0091</td>
<td>0.0660</td>
<td>0.0343</td>
<td>0.0440</td>
<td>0.0109</td>
<td>0.0390</td>
<td>0.0399</td>
<td>0.0138</td>
<td><strong>0.3217</strong></td>
</tr>
<tr>
<td>Yokosuka</td>
<td>0.0094</td>
<td><strong>0.0170</strong></td>
<td>0.0026</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0003</td>
<td>0.0006</td>
<td>0.0006</td>
<td><strong>0.0305</strong></td>
</tr>
<tr>
<td>Yokohama</td>
<td>0.0760</td>
<td>0.0032</td>
<td><strong>0.1161</strong></td>
<td>0.0132</td>
<td>0.0015</td>
<td>0.0000</td>
<td>0.0038</td>
<td>0.0035</td>
<td>0.0021</td>
<td><strong>0.2194</strong></td>
</tr>
<tr>
<td>Kawasaki</td>
<td>0.0431</td>
<td>0.0000</td>
<td>0.0126</td>
<td><strong>0.0191</strong></td>
<td>0.0006</td>
<td>0.0003</td>
<td>0.0018</td>
<td>0.0053</td>
<td>0.0003</td>
<td><strong>0.0830</strong></td>
</tr>
<tr>
<td>Tokyo</td>
<td>0.0444</td>
<td>0.0000</td>
<td>0.0044</td>
<td>0.0012</td>
<td><strong>0.0519</strong></td>
<td>0.0009</td>
<td>0.0029</td>
<td>0.0015</td>
<td>0.0015</td>
<td><strong>0.1076</strong></td>
</tr>
<tr>
<td>Funabashi</td>
<td>0.0091</td>
<td>0.0000</td>
<td>0.0026</td>
<td>0.0003</td>
<td>0.0000</td>
<td><strong>0.0035</strong></td>
<td>0.0026</td>
<td>0.0009</td>
<td>0.0003</td>
<td><strong>0.0194</strong></td>
</tr>
<tr>
<td>Chiba</td>
<td>0.0323</td>
<td>0.0000</td>
<td>0.0035</td>
<td>0.0029</td>
<td>0.0023</td>
<td>0.0044</td>
<td><strong>0.0367</strong></td>
<td>0.0144</td>
<td>0.0009</td>
<td><strong>0.0974</strong></td>
</tr>
<tr>
<td>Anegasaki</td>
<td>0.0413</td>
<td>0.0000</td>
<td>0.0041</td>
<td>0.0038</td>
<td>0.0021</td>
<td>0.0000</td>
<td>0.0117</td>
<td><strong>0.0194</strong></td>
<td>0.0006</td>
<td><strong>0.0830</strong></td>
</tr>
<tr>
<td>Kisarazu</td>
<td>0.0126</td>
<td>0.0006</td>
<td>0.0015</td>
<td>0.0009</td>
<td>0.0006</td>
<td>0.0000</td>
<td>0.0018</td>
<td>0.0000</td>
<td><strong>0.0202</strong></td>
<td><strong>0.0381</strong></td>
</tr>
<tr>
<td>( \sum TAZ\textsubscript{i} )</td>
<td><strong>0.3320</strong></td>
<td>0.0299</td>
<td><strong>0.2135</strong></td>
<td>0.0757</td>
<td><strong>0.1029</strong></td>
<td>0.0199</td>
<td>0.1006</td>
<td>0.0853</td>
<td>0.0402</td>
<td><strong>1.0000</strong></td>
</tr>
</tbody>
</table>
Further research is required to breakdown the traffic within the same TAZ into Sub-TAZs to properly understand the underlying behavior of vessel traffic streams as illustrated in Fig. 4.

Following the above conclusions, a new constraint is introduced to exclude trips within the same TAZ into the tracking algorithm.

Table 4 shows the trip probability distribution of all trips made from $TAZ_i$ to $TAZ_j$ for the time domain covering the period of analysis with the newly introduced constraint excluding traffic between the same TAZs.

Entropy $H$ as defined by Equation 6 and Equation 7 is used to measure the uncertainty associated with the trip distributions within the TAZs, or in other words the likelihood of a trip to occur assuming that all trips are independent from each other. Table 5 shows the Entropy distribution model of all trips made from $TAZ_i$ to $TAZ_j$ for the time domain covering the period of analysis with the newly introduced constraint excluding traffic between the same TAZs.

7. Discussion

Traffic trip distribution with a constraint, excluding trips within the TAZs, shows the conservation of traffic for every Origin and Destination TAZ respectively. In addition to trip combinations excluded by the introduced constraint, 13 other trip combinations (underlined values) are confirmed with certainty to be improbable, namely: Yokosuka/ Kawasaki, Yokosuka/ Tokyo, Yokosuka/ Funabashi, Yokohama/ Funabashi, Kawasaki/ Yokosuka, Tokyo/ Yokosuka, Funabashi/ Yokosuka, Funabashi/ Tokyo, Chiba/ Yokosuka, Anegasaki/ Yokosuka, Anegasaki/ Funabashi, Kisarazu/ Funabashi, and Kisarazu /Anegasaki.

The trip distribution model with constraint, as shown in Table 4, is regarded as a behavioral model with the aim to predict the trip distribution while taking into considerations the unknown parameters.

![Fig. 4 Illustration of possible Sub-TAZs for Yokohama and Tokyo port (Source: Japan Coast Guard website 2014/01/31)](image-url)
related to choosing the ship’s destination. On the other hand, the Entropy distribution model, as shown in Table 5, is not a behavioral model because it determines the trips which are most likely to occur assuming that each trip is happening independently of the others; allowing us to grasp and quantify the uncertainty related destination attribution based on the generated traffic at the Origin TAZ.

Fig. 5 shows that the most attractive port is Yokohama Port and it alone generates and attracts more than 20% of traffic without constraint, and around 15% of traffic with constraint, of the total traffic in Tokyo Bay. Furthermore, Fig. 5 (a) shows that around 70% of traffic without constraint (trip distribution of generated traffic from Tokyo Bay Line TAZ without constraint is 0.3217), and around 60% of traffic with constraint (trip distribution of generated traffic from Tokyo Bay Line TAZ with constraint is 0.3944) originated from within Tokyo Bay.

8. Conclusion
This work introduces a review of the Trip Distribution theory and its fundamentals. The Trip Distribution theory is then used to model vessel traffic

<table>
<thead>
<tr>
<th>TAZ_j</th>
<th>Tokyo Bay Line</th>
<th>Yokosuka</th>
<th>Yokohama</th>
<th>Kawasaki</th>
<th>Tokyo</th>
<th>Funabashi</th>
<th>Chiba</th>
<th>Anegasaki</th>
<th>Kisarazu</th>
<th>( \sum TAZ_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo Bay Line</td>
<td>0.0000</td>
<td>0.0596</td>
<td>0.2320</td>
<td>0.1551</td>
<td>0.1820</td>
<td>0.0682</td>
<td>0.1686</td>
<td>0.1710</td>
<td>0.0816</td>
<td>0.3669</td>
</tr>
<tr>
<td>Yokosuka</td>
<td>0.0611</td>
<td>0.0000</td>
<td>0.0223</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0035</td>
<td>0.0063</td>
<td>0.0063</td>
<td>0.0803</td>
</tr>
<tr>
<td>Yokohama</td>
<td>0.2506</td>
<td>0.0263</td>
<td>0.0000</td>
<td>0.0790</td>
<td>0.0137</td>
<td>0.0000</td>
<td>0.0301</td>
<td>0.0282</td>
<td>0.0182</td>
<td>0.2919</td>
</tr>
<tr>
<td>Kawasaki</td>
<td>0.1797</td>
<td>0.0000</td>
<td>0.0764</td>
<td>0.0000</td>
<td>0.0063</td>
<td>0.0035</td>
<td>0.0160</td>
<td>0.0390</td>
<td>0.0035</td>
<td>0.2278</td>
</tr>
<tr>
<td>Tokyo</td>
<td>0.1805</td>
<td>0.0000</td>
<td>0.0338</td>
<td>0.0114</td>
<td>0.0000</td>
<td>0.0089</td>
<td>0.0243</td>
<td>0.0137</td>
<td>0.0137</td>
<td>0.2103</td>
</tr>
<tr>
<td>Funabashi</td>
<td>0.0596</td>
<td>0.0000</td>
<td>0.0223</td>
<td>0.0035</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0223</td>
<td>0.0089</td>
<td>0.0035</td>
<td>0.0904</td>
</tr>
<tr>
<td>Chiba</td>
<td>0.1488</td>
<td>0.0000</td>
<td>0.0282</td>
<td>0.0243</td>
<td>0.0203</td>
<td>0.0338</td>
<td>0.0000</td>
<td>0.0841</td>
<td>0.0089</td>
<td>0.2212</td>
</tr>
<tr>
<td>Anegasaki</td>
<td>0.1750</td>
<td>0.0000</td>
<td>0.0319</td>
<td>0.0301</td>
<td>0.0182</td>
<td>0.0000</td>
<td>0.0723</td>
<td>0.0000</td>
<td>0.0063</td>
<td>0.2272</td>
</tr>
<tr>
<td>Kisarazu</td>
<td>0.0764</td>
<td>0.0063</td>
<td>0.0137</td>
<td>0.0089</td>
<td>0.0063</td>
<td>0.0000</td>
<td>0.0160</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0987</td>
</tr>
<tr>
<td>( \sum TAZ_i )</td>
<td>0.3655</td>
<td>0.0777</td>
<td>0.2841</td>
<td>0.2123</td>
<td>0.1995</td>
<td>0.0928</td>
<td>0.2278</td>
<td>0.2320</td>
<td>0.1067</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
streams navigating between different TAZs in Tokyo Bay based on AIS data.

In addition to the origin port which is not provided within the AIS data, a data review revealed also that the destination port provided within the AIS data is not accurate and the AIS data entry non-compliance rate is high. Therefore, to solve the destination and origin issues, an algorithm is developed to extract the Origin TAZ and Destination TAZ data from historical ships’ tracks. The algorithm tracks all ships sailing from one TAZ to another, and excludes any other ship that is not provided with an Origin TAZ and/or a Destination TAZ.

A review of all trip probability distributions revealed that trips between the same TAZs --attributable to service boats, leisure passenger ships and the like-- accounts for a trip probability of $\rho = 0.34868$. And since vessel traffic streams navigating with in the same TAZs are not relevant to this analysis, a new constraint was introduced excluding all the traffic within the same TAZ.

The results of the Trip Distributions analysis show that the traffic is conserved between every Origin and Destination TAZ respectively. In addition to trip combinations excluded by the introduced constraint, 13 trip combinations are negated with certainty. Moreover, the results also show that the most attractive port is Yokohama Port and it alone generates and receives more than 20% of traffic without constraint, and around 15% of traffic with constraint; of the total traffic in Tokyo Bay. Furthermore, the results show that around 70% of traffic without constraint, and around 60% of traffic with constraint; is originating from within Tokyo Bay.

The trip distribution model uncertainty --associated with the unknown parameters related to destination attribution-- was evaluated based on the generated traffic at the Origin TAZ.

The results of this work provides an effective tool for evaluating the distribution of vessel traffic streams loads and appraising the level of disorder caused in Tokyo Bay. Furthermore, the model formalizes the trip distribution into a matrix that can be used as a metric for traffic generation and evaluating the fluctuation in traffic and/or TAZ traffic load assignment. Nevertheless, further research is needed to assess the traffic within the same TAZ by breaking-down every TAZ into Sub-TAZs.

References
(7) R. KAR, and S. K. MAZUMDER: Entropy and


