Boundary Estimation of Probabilistic Port Hinterland for Intermodal Freight Transportation Operations

Xinchang Wang

Research Fellow, Center for Maritime Studies, NUS
Ph.D Student, Industrial and Systems Engineering, Georgia Tech

Joint work with Prof. Qiang Meng and Prof. Lixin Miao

December 9, 2013
Outline

1. Introduction
2. Literature Review
3. The Model
4. The Monte Carlo Simulation based Algorithm
5. Numerical Examples
6. Conclusions
Introduction

Importance of Maritime Transport

1. Backbone of international trade and a key engine driving globalization
2. The 80% of global trade by volume and over 70% by value is carried by sea and is handled by ports worldwide.
3. These shares are even higher in most developing countries.

Major players

1. Shippers (a.k.a., freight forwarders, non-vessel operating common carriers)
2. Port operators
3. Ocean carriers, liner shipping companies
4. Inland freight transporters
Introduction

Ocean Networks
River Networks
Road Networks
Rail Networks
Introduction

About APL Container Shipping

1. 311 ports, 15,852 O-D (Port-to-Port) pairs, 133,742 laden TEUs per week, 32,740 empty TEUs per week

2. 76 main ship routes, 260 ships, total shipboard capacity: 727,311 TEUs
# Introduction

Top 30 Container Shipping Lines

**Alphaliner - TOP 100**

Operated fleets as per 02 June 2013

## THE TOP 100 LEAGUE

> The percentage shown on the left of each bar represents the operator's share of the world liner fleet in TEU terms.
> The light coloured bar on the right represents the current orderbook (firm orders).

<table>
<thead>
<tr>
<th>Rnk</th>
<th>Operator</th>
<th>TEU</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>APM-Maersk</td>
<td>2,589,126</td>
<td>15.0%</td>
</tr>
<tr>
<td>2</td>
<td>Mediterranean Shg Co</td>
<td>2,333,413</td>
<td>13.5%</td>
</tr>
<tr>
<td>3</td>
<td>CMA CGM Group</td>
<td>1,486,982</td>
<td>8.6%</td>
</tr>
<tr>
<td>4</td>
<td>Evergreen Line</td>
<td>771,456</td>
<td>4.5%</td>
</tr>
<tr>
<td>5</td>
<td>COSCO Container L</td>
<td>760,398</td>
<td>4.4%</td>
</tr>
<tr>
<td>6</td>
<td>Hapag-Lloyd</td>
<td>703,032</td>
<td>4.1%</td>
</tr>
<tr>
<td>7</td>
<td>APL</td>
<td>623,523</td>
<td>3.6%</td>
</tr>
<tr>
<td>8</td>
<td>Hanjin Shipping</td>
<td>623,002</td>
<td>3.6%</td>
</tr>
<tr>
<td>9</td>
<td>COSCL</td>
<td>602,477</td>
<td>3.5%</td>
</tr>
<tr>
<td>10</td>
<td>MOL</td>
<td>631,206</td>
<td>3.1%</td>
</tr>
<tr>
<td>11</td>
<td>OOCL</td>
<td>461,232</td>
<td>2.7%</td>
</tr>
<tr>
<td>12</td>
<td>NYK Line</td>
<td>432,003</td>
<td>2.5%</td>
</tr>
<tr>
<td>13</td>
<td>Hamburg Sud Group</td>
<td>414,348</td>
<td>2.4%</td>
</tr>
<tr>
<td>14</td>
<td>Yang Ming Marine Transport Corp.</td>
<td>379,914</td>
<td>2.2%</td>
</tr>
<tr>
<td>15</td>
<td>K Line</td>
<td>361,616</td>
<td>2.1%</td>
</tr>
<tr>
<td>16</td>
<td>PIL (Pacific Int. Line)</td>
<td>354,138</td>
<td>2.1%</td>
</tr>
<tr>
<td>17</td>
<td>Hyundai M. M.</td>
<td>338,923</td>
<td>2.0%</td>
</tr>
<tr>
<td>18</td>
<td>Zim</td>
<td>337,515</td>
<td>2.0%</td>
</tr>
<tr>
<td>19</td>
<td>UASC</td>
<td>263,116</td>
<td>1.5%</td>
</tr>
<tr>
<td>20</td>
<td>CMA Group</td>
<td>250,490</td>
<td>1.5%</td>
</tr>
<tr>
<td>21</td>
<td>Wan Hai Lines</td>
<td>165,386</td>
<td>1.0%</td>
</tr>
<tr>
<td>22</td>
<td>HDS Lines</td>
<td>66,320</td>
<td>0.6%</td>
</tr>
<tr>
<td>23</td>
<td>X-Press Feeders Group</td>
<td>76,968</td>
<td>0.4%</td>
</tr>
<tr>
<td>24</td>
<td>Niedutch</td>
<td>70,640</td>
<td>0.4%</td>
</tr>
<tr>
<td>25</td>
<td>TS Lines</td>
<td>67,202</td>
<td>0.4%</td>
</tr>
<tr>
<td>26</td>
<td>SITC</td>
<td>66,163</td>
<td>0.4%</td>
</tr>
<tr>
<td>27</td>
<td>KMTC</td>
<td>58,254</td>
<td>0.3%</td>
</tr>
<tr>
<td>28</td>
<td>RCL (Regional Container L.)</td>
<td>51,194</td>
<td>0.3%</td>
</tr>
<tr>
<td>29</td>
<td>CCNI</td>
<td>43,679</td>
<td>0.3%</td>
</tr>
<tr>
<td>30</td>
<td>UniFeeder</td>
<td>42,650</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

All information above is given as guidance only and in good faith without guarantee

© Alphaliner 1999-2013
The concept of destination-based port hinterland

The destination-based port hinterland of a port $P$ refers to an area served by the port, over which shippers transport their cargo to a specific destination $d$ via $P$. 
**Introduction**

**Definition (Probabilistic port hinterland)**
An area surrounding the port, over which shippers choose the port to transport their goods to a given destination \(d\) with probability within \([\alpha, 1]\), where \(\alpha \in [0, 1]\).

**Definition (\(\alpha\)–Boundary of probabilistic port hinterland)**
The \(\alpha\)–Boundary of port hinterland refers to a curve/line, on which all shippers will choose to use the port with the same probability \(\alpha \in [0, 1]\).
Introduction

**The objective:**

The objective of our work is to develop an analytical approach to estimating the destination-based $\alpha$—boundaries of probabilistic hinterland of a specific port in the setting of intermodal freight transportation operations.
Qualitative Analysis of Port Hinterland:

1. Van Cleef, E. (1941): Hinterland and umbland
**Port choice of freight forwarders:**


**Comments:**

1. Research methodologies: questionnaire survey and case studies

2. Impact factors: port service, transportation cost, transportation time, hinterland connectivity

3. Implications: port hinterland are affected by two primary factors: transportation cost and time, namely, utility
Literature Review

**HINTERLAND ESTIMATION APPROACHES:**

The Model

1. \( o \): the origin with coordinates \( o = (x, y) \) on the \( x - y \) plane
2. \( d \): the destination
3. \( R(x, y) \): the set of intermodal routes connecting \( (o, d) \), \( l \in R(x, y) \)
4. \( R_P(x, y) \): the set of intermodal routes traversing through \( P \).
5. \( \bar{R}_P(x, y) \): \( = R(x, y) \setminus \bar{R}_P(x, y) \).
6. \( a \in A_l \): links; \( n_1 \in T_l \): transfer nodes; \( n_2 \in N_l \): regular nodes.
The Model

We assume independent and Gaussian distributed utilities (VOT \times \text{time-cost}) over links and at transfer nodes, i.e., for each \( l \in R(x, y) \)

\[
\begin{align*}
    u_a(x, y) &= \eta_a(x, y) + \xi_a(x, y), \quad \forall a \in A_l, \\
    u_n(x, y) &= \lambda_n(x, y) + \omega_n(x, y), \quad \forall n \in T_l,
\end{align*}
\]

where \( \eta_a(x, y) \) and \( \lambda_n(x, y) \) are \((x, y)\)-dependent expected utilities and \( \xi_a(x, y) \) and \( \omega_n(x, y) \) are Gaussian random errors with zero means.

Thus, the utility of any intermodal route \( l \in R(x, y) \) can be represented by,

\[
U_l = \sum_{a \in A_l} u_a(x, y) + \sum_{n \in T_l} u_n(x, y).
\]

Let \( U_{xy} = (U_l, l \in R(x, y)) \).
The Model

Travel distance

\[ V \]

\[ U_l, l \in R(x, y) \]

\[ o = (x, y) \]

\[ n_1 \]

\[ n_2 \]

\[ d \]

\[ \eta_{n_2d} \]

\[ \lambda_{n_1} \]

\[ \eta_{on_1} \]
The Model

All intermodal route utilities form a Gaussian vector with

1. the expected utility for each route \( l \in R(x, y) \),

\[
V_l := \mathbb{E}(U_l) = \sum_{a \in A_l} \eta_a(x, y) + \sum_{n \in T_l} \lambda_n(x, y),
\]

2. the variance for each route \( l \in R(x, y) \),

\[
\delta_l^2 := \text{Var}(U_l) = \sum_{a \in A_l} \text{Var}(\xi_a(x, y)) + \sum_{n \in T_l} \text{Var}(\omega_n(x, y)),
\]

3. the covariance between any two routes \( k, l \in R(x, y) \),

\[
\text{Cov}(U_l, U_k) = \sum_{a \in A_l \cap A_k} \text{Var}(\xi_a(x, y)) + \sum_{n \in T_l \cap T_k} \text{Var}(\omega_n(x, y)),
\]

The Model

THE PROBABILISTIC HINTERLAND:

\[ \Psi_P(\alpha) := \{(x, y) : p(x, y) \geq \alpha\}, \]

where,

\[ p(x, y) := \sum_{l \in R_P(x, y)} \mathbb{P} \left( l \in \arg \max_{k \in R(x, y)} \{U_k\} \right) = \sum_{l \in R_P(x, y)} p_l \]

If \( \Sigma_{xy} := (\text{Cov}(U_l, U_k))_{l, k \in R(x, y)} > 0 \) (which is not quite necessarily strict),

\[ p_l := \int_{\mathbf{z} \leq z_l e} \frac{\exp \left\{ -\frac{1}{2} (\mathbf{z} - \mathbf{V}_{xy})^\top \Sigma_{xy}^{-1} (\mathbf{z} - \mathbf{V}_{xy}) \right\}}{\sqrt{(2\pi)^{I_{xy}} \det(\Sigma_{xy})}} d\mathbf{z}, \]

\[ \mathbf{V}_{xy} := (V_l, l \in R(x, y)), \]

\[ I_{xy} := \text{Card}(R(x, y)). \]
The $\alpha$–boundary is represented by

$$\partial \Psi_P(\alpha) = \{(x, y) : p(x, y) = \alpha\},$$

where $\alpha \in [0, 1]$. 

\[
\Psi_P(\alpha) = \{(x, y) : p(x, y) \geq \alpha\}
\]
The Monte Carlo Simulation based Algorithm

**The algorithm**

**Step 1.** Discretization. Generate a grid. Let \( \hat{\Omega} \) denote the set of intersections on the grid and take them as our origins \((x, y)\).

**Step 2.** Calculate the estimate \( \hat{p}(x, y) \). For each \((x, y) \in \hat{\Omega} \) on the grid,

1. compute \( V_{xy} \) and \( \Sigma_{xy} \).
2. generate \( N \) i.i.d. samples \( \sim \) Gaussian\((V_{xy}, \Sigma_{xy})\), \( \{\hat{U}_{xy}^i\}_{i=1}^N \).
3. do the counting.

\[
\hat{p}(x, y) = \sum_{l \in R_P(x, y)} \frac{K_l}{N},
\]

where

\[
K_l = \sum_{i=1}^{N} 1_{[l \in \arg\max_k \{ \hat{U}_{k} \} \cap R(x, y)]}.
\]

\( R_P(x, y) \) is the region surrounding \( (x, y) \).
The Monte Carlo Simulation based Algorithm

The algorithm

**Step 3.** Determine the points on the $\alpha-$boundary curve with a small error $\varepsilon > 0$, e.g., $\alpha = .35$ and $\varepsilon = .0001$.

$$\partial \hat{\Psi}(\alpha, \varepsilon) = \{(x, y) \in \hat{\Omega} : \hat{p}(x, y) \in [\alpha - \varepsilon, \alpha + \varepsilon]\}.$$
The algorithm

**Step 3.** Determine the points on the \( \alpha \)-boundary curve with a small error \( \varepsilon > 0 \), e.g., \( \alpha = .35 \) and \( \varepsilon = .0001 \).

\[
\partial \hat{\Psi}(\alpha, \varepsilon) = \{(x, y) \in \hat{\Omega} : \hat{p}(x, y) \in [\alpha - \varepsilon, \alpha + \varepsilon]\}.
\]

**Step 4.** \( \alpha \)-boundary curve fitting.

- Cluster analysis
The algorithm

**Step 3.** Determine the points on the $\alpha$–boundary curve with a small error $\varepsilon > 0$, e.g., $\alpha = .35$ and $\varepsilon = .0001$.

$$\partial \hat{\Psi}(\alpha, \varepsilon) = \{ (x, y) \in \hat{\Omega} : \hat{p}(x, y) \in [\alpha - \varepsilon, \alpha + \varepsilon] \}.$$ 

**Step 4.** $\alpha$–boundary curve fitting.

1. Cluster analysis
2. Polynomial curve fitting.
The Monte Carlo Simulation based Algorithm

Issue 1: how do we determine the number of clusters?

Elbow point: $k = 6$
Issue 2: how do we trust our approximation of $\hat{p}(x, y)$?

It follows from the Central Limit Theorem that the confidence interval (w.r.t. .05 significance level) is,

$$\hat{p}(x, y) \pm 1.96 \sqrt{\frac{1 - \sum_{l \in R_P(x, y)} p_l}{N} \sum_{l \in R_P(x, y)} p_l}.$$ 

It follows that

$$\max_{p_l : l \in R_P(x, y)} \left( 1 - \sum_{l \in R_P(x, y)} p_l \right) \sum_{l \in R_P(x, y)} p_l = \frac{1}{4}$$

where the maximality is achieved at $p_l^* = \frac{1}{2\text{Card}(R_P(x, y))}$ for all $l \in R_P(x, y)$. 

The Monte Carlo Simulation based Algorithm
Issue 3: what is a reasonably good sample size?

one may wonder how much computational effort is required in order to guarantee the confidence interval to have length no greater than $2\gamma > 0$. The lower bound on the sample size turns out to be

$$N \geq \left(\frac{1.96}{2\gamma}\right)^2$$

For example, $\gamma = .0001$ gives that $N \geq 9800$. 

The example network of three intermodal routes

\[ E = (-2171, 0), \quad C = (2171, 950) \text{ and } A = (2171, 0) \]
Numerical Examples

**Data: Transportation Cost**

1. Truck transportation: the fare for transporting about 200 km is estimated at 300 USD (USCC 2006).

2. Long-haul rail transportation: Wang et al. (2009) calibrate the following cost-distance function for rail transportation,

\[ c(l) = 268 + 0.267l, \quad r^2 = 0.717. \]  

(1)
Numerical Examples

**Data: transportation cost**

1. Border crossing cost: UNESCAP (2003) reports that the cost incurred in the border crossing process between China and Mongolia is about 293 USD.

2. Port operation costs: see Table 1

3. Maritime transportation cost: see Table 1.
Table 1: Transportation Costs of the Three Intermodal Routes

<table>
<thead>
<tr>
<th>Route 1</th>
<th>E(Cost)</th>
<th>Route 2</th>
<th>E(Cost)</th>
<th>Route 3</th>
<th>E(Cost)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oH₁</td>
<td>300</td>
<td>oH₂</td>
<td>300</td>
<td>oH₃</td>
<td>300</td>
</tr>
<tr>
<td>H₁</td>
<td>268</td>
<td>H₂</td>
<td>268</td>
<td>H₃</td>
<td>268</td>
</tr>
<tr>
<td>H₁E</td>
<td>(1)</td>
<td>H₂A</td>
<td>(1)</td>
<td>H₃A</td>
<td>(1)</td>
</tr>
<tr>
<td>B</td>
<td>293</td>
<td>A</td>
<td>387</td>
<td>C</td>
<td>80</td>
</tr>
<tr>
<td>Ed</td>
<td>300</td>
<td>AE</td>
<td>850</td>
<td>CE</td>
<td>1000</td>
</tr>
<tr>
<td>Ed</td>
<td>300</td>
<td>Ed</td>
<td>300</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Numerical Examples

Some more considerations:

1. The variance of links are assumed to be \((0.2\mathbb{E}(\text{Cost}))^2\).

2. The covariance of two routes is equal to the sum of the variances of common links and transfer nodes shared by these two routes.

3. When the port \(A\) or \(C\) is less than 200 km away from the origin \(o = (x, y)\), only trucking service is used for domestic land transportation.
Case 1: Route 1 and Route 2

We have closed-form expressions by using the c.d.f. of standard normal random variables, \( \Phi(\cdot) \).

Within 200 km away from Shanghai port \( A \) (the circular area centered \( A \)),

\[
\partial \Psi_A(\alpha) = \left\{ (x, y) : \frac{C_{H1E} - 676}{\sqrt{48400 + (0.2C_{HE})^2}} = \Phi^{-1}(\alpha) \right\},
\]

where \( C_{H1E} = 0.267(\| (x, y) - (-2171, 0) \|_2 - 200) \).
The example network of three intermodal routes

- Maritime
- Land transport
- The origin
- The destination
- Rail-truck terminals
- The port
- The border

Origin: Shanghai port A

Destination: Tianjin port C

Route 1: Yangon port E → d → Route 1
Route 2: Route 2
Route 3: Route 3

China

Rail-truck terminals

Shanghai port A

Tianjin port C

Border

Maritime transport

Land transport

The origin

The destination

Rail-truck terminals

The port

The border
Case 1: Route 1 and Route 2

More than 200 km away from $A$ (outside the circular area),

$$\partial\Psi_A(\alpha) = \left\{ (x, y) : \frac{C_{H1E} - C_{H2A} - 944}{\sqrt{51271 + (0.2C_{HE})^2 + (0.2C_{H2A})^2}} = \Phi^{-1}(\alpha) \right\},$$

where $C_{H2A} = 0.267(\| (x, y) - (2171, 0) \|_2 - 200)$. 

**Numerical Examples**

### Case 1: Route 1 and Route 2

![Graph showing simulation results vs. closed-form expressions](image)

Simulation results v.s. closed-form expressions

- **Fitting curves**
- **Analytical curves**
- **Simulation points**

**Fitting curves**

\[
\begin{align*}
\alpha &= 0.1 \\
\alpha &= 0.2 \\
\alpha &= 0.5
\end{align*}
\]

- \( f_{0.1}(y) = 2.18 \times 10^{-10} y^3 + 1.01 \times 10^{-4} y^2 + 1.25 \times 10^{-3} y + 1093, R^2 = 0.985 \)
- \( f_{0.2}(y) = 1.23 \times 10^{-9} y^3 + 1.60 \times 10^{-4} y^2 - 3.51 \times 10^{-4} y + 1322, R^2 = 0.991 \)
- \( f_{0.5}(y) = 4.79 \times 10^{-9} y^3 + 4.38 \times 10^{-4} y^2 - 7.97 \times 10^{-4} y + 177, R^2 = 0.995 \)
- \( f_{0.90}(y) = 1.03 \times 10^{-8} y^3 - 1.58 \times 10^{-3} y^2 - 0.018 y + 2047, R^2 = 0.508 \)

Shanghai port A
**Case 2: three competing routes, Route 1, Route 2 and Route 3**

\[
f_{0.015}(y) = -1.26 \times 10^{-4} y^3 + 0.35 y^2 - 320.48 y + 99411, R^2 = 0.157
\]

\[
f_{0.80}(y) = -1.86 \times 10^{-5} y^3 + 3.25 \times 10^{-3} y^2 + 2.20 y + 2136, R^2 = 0.929
\]

\[
f_{0.10}(y) = 9.86 \times 10^{-8} y^3 + 5.04 \times 10^{-4} y^2 + 0.367 y + 1180, R^2 = 0.934
\]

\[
f_{0.20}(y) = 7.58 \times 10^{-7} y^3 + 1.62 \times 10^{-3} y^2 + 0.54 y + 1376, R^2 = 0.937
\]
Numerical Examples

Case 2: three competing routes, Route 1, Route 2 and Route 3

Fitting curves corresponding to probability 0.35

Simulation results with cluster analysis
Numerical Examples

The number of clusters
Elbow point: \( k = 6 \)

Percentage of variance explained: Between-group variance
Total variance

The number of clusters
Conclusions

1. We defined probabilistic port hinterland boundary for intermodal freight transportation operations.
2. We proposed an algorithm to identify $\alpha$—boundaries of probabilistic port hinterland.
3. We demonstrated our approach by numerical examples.
Thank you all!

Questions?