SELECTED PROCEEDINGS

Boundary Estimation of Probabilistic Port Hinterland for Intermodal Freight Transportation Operations

QIANG MENG,A DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING, NATIONAL UNIVERSITY OF SINGAPORE, SINGAPORE

XINCHANG WANG, RESEARCH CENTER FOR MODERN LOGISTICS, GRADUATE SCHOOL AT SHENZHEN, TSINGHUA UNIVERSITY, CHINA

LIXIN MIAO, RESEARCH CENTER FOR MODERN LOGISTICS, GRADUATE SCHOOL AT SHENZHEN, TSINGHUA UNIVERSITY, CHINA.

This is an abridged version of the paper presented at the conference. The full version is being submitted elsewhere. Details on the full paper can be obtained from the author.

ISBN: 978-85-285-0232-9

13th World Conference on Transport Research

www.wctr2013rio.com

15-18 JULY 2013

Rio de Janeiro, Brazil

Boundary Estimation of Probabilistic Port Hinterland for Intermodal Freight Transportation Operations

Qiang Meng [a], Xinchang Wang [a] and Lixin Miao[b]
[a] *Department of Civil and Environmental Engineering, National University of Singapore, Singapore 117576*
[b] *Research Center for Modern Logistics, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China*

ABSTRACT

This paper develops a discreet choice theory based approach to estimating boundaries of the probabilistic hinterland of a port for intermodal freight transportation operations. We first define the probabilistic hinterland of a port, and then derive its mathematical formulation by assuming that transportation costs or times of all available intermodal routes from an origin to a destination are multivariate normally distributed. We develop a Monte Carlo simulation based algorithm to graphically depict the port hinterland boundary with respect to a certain probability value. The algorithm comprises an interesting boundary curve fitting procedure which is realized by using cluster analysis and the least-square estimation method. In addition, a lower bound on the sample size required by the Monte Carlo simulation is derived. Finally, two illustrative examples are given to show the effectiveness and applicability of the proposed approach.

Keywords: Intermodal transportation, Port hinterland, Probit-based route choice, Monte Carlo simulation, Cluster analysis.

1 INTRODUCTION

The recent changes in the world economy due to market globalization have triggered a substantial increase in demand for international seaborne trade and intermodal freight transportation services. As reported by UNCTD (2007), the seaborne trade accounts for a major proportion of world trade. A port is considered to be a critical gateway or transshipment hub, which provides infrastructure facilities for cargo movements between inland and maritime modes, and it is play an important role in supply chain (Tongzon et al, 2009). Mega ports such as Singapore, Hong Kong and Shanghai ports have been playing a key role in determining the economic well-being of the areas served by them. The area served by a particular port, over which shippers transport their goods to a specific destination via the port, is referred to as the port's hinterland with respect to the destination. Dedicated port hinterlands are deemed to facilitate smooth intermodal movements of goods and to ensure goods reach their final destinations quicker and more cheaply (UNESCAP 2005). With the help of flexible sea-land linkage, port hinterland can be utilized as a dry port or warehouse to save the valuable waterfront area in a port by means of intermodal transportation operations (Parolaand and

Sciomachen, 2005; Xie 2009; Wang and Yun, 2011). Thus, port hinterland can be thought of as a performance indicator to reflect the competitiveness of a port, since it represents the attraction of a port to shippers from a spatial point of view.

The boundary estimation of a port's hinterland is concerned with identifying the boundary of a logistics zone, over which shippers incline to select a specific port to transship their goods. In actual, the port selection is a companion result of intermodal route choice of shippers. Shippers choose the port by choosing intermodal routes traversing though it. To identity the hinterland boundary, in addition to the level of service provided by the port itself, available logistics services including truck, maritime and rail transportation should also be taken into account. These services may have either positive or negative impact on the port hinterland, given their travel time and cost effectiveness. For example, the hinterland of the Pusan port of South Korea covers not only South Korea, but also parts of Northern China and even Japan because cargo drawn from these areas can be transported to Pusan port by fast and economical feeder services (Yeo et al. 2008).

Given a shipper with a specific location (origin) who can access a specific port and desires goods to be transported to a destination, there are usually several intermodal routes available between the origin and destination. An intermodal route can involve multiple modes of transportation and several transfer nodes. A transfer node functions as an interacting point that switches goods between multiple modes. Examples include rail-truck terminals, ports, and border crossing terminals between two neighbored countries. According to Yeo et al. (2008), transportation cost and time are two crucial factors in determining a shipper's port selection. The port selection, however, results from the route choice of the shipper from a set of available intermodal routes. In practice, the transportation cost or time of a particular intermodal route may vary based on intermodal carriers and transportation market price. It can be appropriately represented by a random variable. To describe the decision behaviors of shippers faced with route choice, we assume that a shipper would choose the route with the minimum transportation cost or time to transport his/her goods from a set of available ones. Under this assumption, shippers will choose a port with some probability instead of certainly selecting it. We can thus define the probabilistic hinterland of a port as: an area surrounding the port, over which shippers choose the port to transport their goods to a given destination with a certain probability value in range $[\alpha,1]$, where $\alpha \in [0,1]$. The boundary of the probabilistic port hinterland with respect to α represents a line, on which all shippers will choose the port with probability α. Boundary estimation of a port's hinterland can benefit a broad range of communities. For local governments who have ports as the main engine of economic development, port hinterland analysis can help in designing an efficient and accessible hinterland network to attract more cargo flows, so as to facilitate the development of local economies. By analyzing port hinterland, a port operator can identify key factors affecting the size and extent of port market area. One can refer to Wang et al. (2009) and Wang and Meng (2010) for analysis of landbridge on the market shares of Asian ports. Moreover, a shipper can choose a proper port to handle his/her goods after analyzing the hinterlands of available ports.

This paper aims to develop a modeling approach to estimating the boundaries of the probabilistic hinterland of a specific port in the context of intermodal freight transportation operations. In this paper we focus on container transportation, since containerized cargo is currently the main type of the goods handled by intermodal freight transportation. In addition, without loss of the generality, our approach is developed based on transportation cost. However, it must be pointed out that the proposed approach can also be applied to the scenarios where bulk transportation, transportation time, reliability and other port selection criteria are concerned.

1.1 Past Related Studies

Van Cleef (1941) seems to be the first scholar who defines hinterland for a trade center. He described the hinterland as an area adjacent to the trade center, with which economic and some culture activities are focused largely on the primary center. This notation of hinterland is quite similar to the hinterland of a port as far as the port is considered as a trade center. Many studies have been conducted to qualitatively analyze port hinterland related problems in transport geography field, such as the growth and coverage of a port, the function of a port in regionalization and globalization, and port-city relations and industrial changes (Lee et al. 2008). All these studies emphasized that intermodal freight transportation services could generate important impact on port hinterland. Hoare (1986) also noticed the potential impact of intermodalism on port hinterland. He suggested that the above-mentioned notion of port hinterland should be revisited to be adapted to the changes in the context of intermodalism. Van Klink and Van den Berg (1998) proposed a definition of port hinterland that takes into intermodal freight transportation operations into account. By their definition, port hinterland is described as the continental area of origin and destination of freight traffic flows through a port. To further examine the spatial and functional nexus that port hinterland has become, Notteboom and Rodrigue (2007) defined physical port hinterland by considering transportation supply from both unimodal and intermodal points of view. These scholars, however, have not developed available methodologies that can quantitatively identify or estimate the hinterland of a port.

Port hinterland estimation is closely related to port selection of shippers. Port selection criteria are direct and important tools that make us understand the behavior of shippers faced with port selection. Numerous studies have been published to elaborate on this issue. One may refer to Yeo et al. (2008) and Tongzon (2008) for detailed discussions. According to their survey papers, various criteria such as port service, transport cost and efficiency, hinterland connectivity have been identified as important factors influencing a shipper’s port selection. The methods to analyze the criteria include questionnaire surveys (Chang et al. 2008), case studies (Garrido and Leva 2004; De Langen 2007) and a game theoretical model (Zan 1999). Malchow and Kanafani (2001) deemed that the oceanic and inland distances are the most significant factors impacting a shipper’s port selection. Yeo et al. (2008) also found that the most competitive ports rely on their efficient hinterland logistics systems. It is implied that transportation cost and time are two primary factors that ought to be concerned in analyzing port selection of shippers. Port hinterland should thus be estimated by taking into consideration of transportation cost and time.

Some quantitative analysis methods have been proposed to estimate the market area of a trade center or transfer terminal, which would inspire us to estimate port hinterland. Fetter (1924) and Hyson and Hyson (1950) derived the hyperbola boundary for the market area of a trade center by assuming that the unit transportation cost and market price for shippers are constant. Niérat (1997) employed a similar method to estimate the market area of an intermodal rail-road terminal, and analyzed the key factors influencing the size of it. O'Kelly and Miller (1989) utilized a gravity model to estimate the probability-based boundaries of the market area of a firm. This model simply adopts the reciprocal of travel distance as deterrence function and does not involve the uncertainties in transportation cost or time. Discrete choice analysis methods have been used to investigate the port selection of shippers. Malchow and Kanafani (2004) applied logit-based disaggregate model to analyze the route choice as well as port selection of shippers. Tang et al. (2011) developed a network-based integrated choice evaluation model to identify port choice criteria of liner shipping companies. The model was believed to enhance the multinomial logit model by integrating network attributes. However, the logit-based modes can perform well only when the correlation between choice alternatives can be ignored or the presumed structure of covariance matrix is adopted. These drawbacks restrict the application of logit-based models in handling choice of freely correlated alternatives, which is however easily seen in intermodal route choice. Meng and Wang (2010) developed a probit-based method to estimate probabilistic port hinterland from a network point of view. The method identifies port hinterland as an area corresponding to a probability range and does not move forward to estimate the boundary corresponding to a certain probability value. In addition, the computational effort needed in running simulation to obtain a targeted accuracy is not specified.

1.2 Contributions of the Study

This study actually extends the work of Meng and Wang (2010) to the estimation of the boundaries of the probabilistic hinterland of a specific port for intermodal freight transportation operations. The contributions of the study are four-fold:

(i) We propose a new definition of probabilistic port hinterland which contributes an alternative perspective to envision the competiveness of a port.

(ii) With the assumption that transportation costs of all available intermodal routes from an origin to a destination are multivariate normally distributed, we develop a mathematical model to formulate the probabilistic hinterland of a port and its boundaries with respect to certain probability values.

(iii) A Monte Carlo simulation based algorithm is designed to determine and graphically show the boundaries of the port hinterland. The algorithm comprises an interesting boundary curve fitting procedure which is realized by using cluster analysis and the least-square estimation method. The algorithm is able to estimate discontinuous boundary curves, which result from the piecewise linear characteristic of intermodal routes.

(iv) A lower bound on the sample size needed in the simulation to achieve the significance level of 95% is derived when approximating the probability of shippers selecting the specific port.

The remainder of this paper is organized as follows. Section 2 describes notations, assumptions and problem statement. Section 3 develops a mathematical model to formulate the probabilistic hinterland of a specific port and its boundaries. The Monte Carlo simulation based algorithm is discussed in Section 4. Two illustrative examples are given in Section 5 to show the application of the model and algorithm. Conclusions are drawn in Section 6.

2 NOTATIONS, ASSUMPTIONS AND PROBLEM STATEMENT

To represent the hinterland boundaries of a specific port *P*, we use an *x*-*y* plane coordinate system as shown in Fig. 1, where *P* is located on the *x*-axis. Given a point (*x*, *y*) as an origin surrounding P in the *x*-*y* plane, the shippers located around (*x*, *y*) intend to transport containers from the origin to a specified destination S through available intermodal freight transportation routes. Let *R*(*x*, *y*) be the set of all these available routes. *R*(*x*, *y*) can be classified into two exclusive subsets: $R_P(x,y)$ and $\overline{R}_P(x,y)$, where $R_P(x,y)$ consists of the routes traversing through *P* and $\overline{R}_P(x,y)$ contains the routes that do not pass by *P*. We thus have,

$$R(x,y) = R_P(x,y) \cup \overline{R}_P(x,y). \tag{1}$$

One may notice that intermodal container transportation is generally long-haul transportation that may traverse across several cities, countries or even continents. The number of all possible intermodal routes is often manageable and can be enumerated in practice.

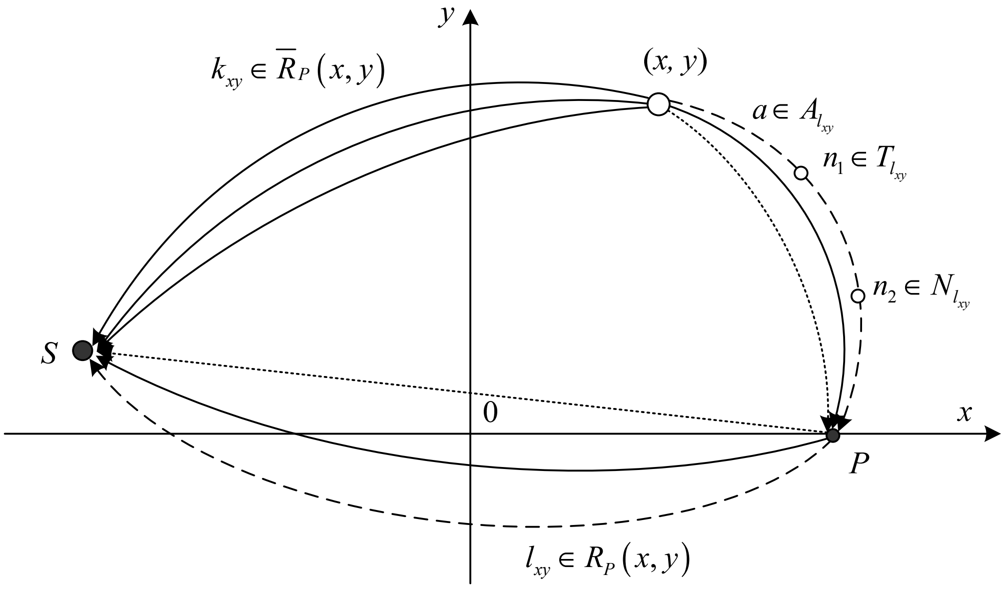


Figure 1 An *x*-*y* coordinate system involving intermodal routes

Let l_{xy} denote an intermodal route from (x,y) to S, i.e. $l_{xy} \in R(x,y)$. It may consist of a sequence of nodes (i.e., points) as well as the links between two consecutive nodes in the *x*-*y* plane. A node can either represent a transfer node where containers are switched between different modes or a location where the handling of containers does not happen. The latter is referred to as regular node. One example of the regular node is the interaction of two roadways. Let $T_{l_{xy}}$ and $N_{l_{xy}}$ be the sets of all transfer and regular nodes on route l_{xy}, respectively. We restrict that each link on route l_{xy} possesses only one mode such as truck, rail or maritime. If more than one mode exists between two consecutive nodes, each mode should be represented by one individual link.

Let $A_{l_{xy}}$ be the set of links on route l_{xy}. For each link $a \in A_{l_{xy}}$, let $C_a(x,y)$ represent the cost incurred in transporting one TEU (Twenty Equivalent Unit) over the link. For each transfer node $n \in T_{l_{xy}}$, let $C_n(x,y)$ represent the cost of transshipping one TEU at the node. To capture the uncertainties in the costs, we assume $C_a(x,y)$ and $C_n(x,y)$ to be normally distributed random variables. The assumption is made because the cost incurred over a link or at a transfer node generally varies based on transportation companies and the transportation market price. The distribution followed by two random variables is, of course, not necessarily normal, and depends on the calibration and verification of historical data. However, since normal distribution can be used to describe, at least approximately, any random variable that trends to cluster around its mean value, it has been extensively employed in practical applications (Patel and read, 1996). Mathematically, the two random variables can be represented as,

$$C_a(x,y) = \eta_a(x,y) + \xi_a(x,y), a \in A_{l_{xy}} \tag{2}$$

$$C_n(x,y) = \lambda_n(x,y) + \omega_n(x,y), n \in T_{l_{xy}} \tag{3}$$

where $\eta_a(x,y)$ and $\lambda_n(x,y)$ are the expected values of $C_a(x,y)$ and $C_n(x,y)$, respectively, given origin (x,y), and $\xi_a(x,y)$ and $\omega_n(x,y)$ are two zero-mean random error terms that reflect the variations on the costs incurred over link $a \in A_{l_{xy}}$ and at transfer node $n \in T_{l_{xy}}$, respectively. The cost along route l_{xy} can thus be represented as,

$$U_{l_{xy}} = \sum_{a \in A_{l_{xy}}} C_a(x,y) + \sum_{n \in T_{l_{xy}}} C_n(x,y) \tag{4}$$

We assume that $C_a(x,y)$ and $C_n(x,y)$ are statistically independent for each $a \in A_{l_{xy}}$ and $n \in T_{l_{xy}}$. It therefore follows that $U_{l_{xy}}$ is also a normally distributed random variable.

As a reasonable assumption, given destination *S*, shippers around location (*x*, *y*) exhibit the same behavior in route choice, and they would choose the route with the minimum transportation cost from *R*(*x*, *y*) to transport their containers. The boundary estimation of the probabilistic hinterland of a specific port *P* aims to identify the boundaries of an area surrounding *P*. On each of the boundaries, shippers transport containers to *S* via the intermodal routes traversing though *P* with a certain probability value.

3 MODEL DEVELOPMENT FOR THE BOUNDARIES OF PROBABILISTIC PORT HINTERLAND

3.1 Piecewise Linear Characteristic Of Intermodal Routes

Let $V_{l_{xy}}$ and $\delta_{l_{xy}}$ be the expected value and variable of $U_{l_{xy}}$, where $l_{xy} \in R_{xy}$, respectively. They can be represented as,

$$V_{l_{xy}} = \mathrm{E}\left(U_{l_{xy}}\right) = \sum_{a \in A_{l_{xy}}} \eta_a(x,y) + \sum_{n \in T_{l_{xy}}} \lambda_n(x,y) \tag{5}$$

$$\delta_{l_{xy}} = \mathrm{var}\left(U_{l_{xy}}\right) = \sum_{a \in A_{l_{xy}}} \mathrm{var}\left[\xi_a(x,y)\right] + \sum_{n \in T_{l_{xy}}} \mathrm{var}\left[\omega_n(x,y)\right] \tag{6}$$

Eqns. (5) and (6) indicate that the expected value or variance of an intermodal route is summed by that of each link and transfer node along the route. For intermodal freight transportation operations, the expected value of the transportation cost over a link can be considered as a linear function of travel distance, and the expected transshipment cost at a transfer node can be considered as a constant value. These considerations give rise to a piecewise linearly shaped expected cost of intermodal routes.

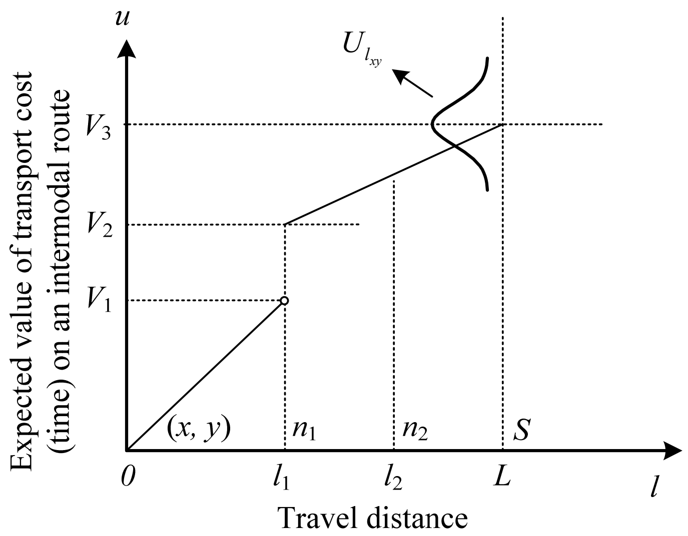


Figure 2 The piecewise linear function of an intermodal route

As shown in Fig.2, the expected values of transport costs from origin (x, y) to transfer node n_1 and from n_1 to destination S are linearly proportional to travel distance. The expected value of transshipment cost at n_1 is a constant value. No transshipment cost is incurred at regular node $n_2 \in N_{l_{xy}}$. As indicated by the figure, V_1 and (V_3-V_2) are the expected values of transportation costs incurred over two links: from (x, y) to n_1 and from n_1 to S, respectively; (V_2-V_1) is the expected transshipment cost expended at n_1. The piecewise linear cost function reflects the unique characteristic of intermodal freight transportation that consists of both container handing at transfer nodes and transportation over links. This characteristic has an important impact on the shape of the boundaries of port hinterland.

3.2 Covariance Matrix

Based on eqn. (4), the covariance of two routes l_{xy}, $k_{xy} \in R(x,y)$ in terms of cost is given by,

$$\mathrm{cov}\left(U_{l_{xy}}, U_{k_{xy}}\right) = \sum_{a \in A_{l_{xy}} \bigcap A_{k_{xy}}} \mathrm{var}\left[\xi_a(x,y)\right] + \sum_{n \in T_{l_{xy}} \bigcap T_{k_{xy}}} \mathrm{var}\left[\omega_n(x,y)\right] \tag{7}$$

Eqn. (7) implies the covariance between two intermodal routes is summed by the variances of the common links and transfer nodes shared by them. The covariance matrix of all the routes in set $R(x,y)$ can be expressed by

$$\Sigma_{xy} = \left[\mathrm{cov}\left(U_{l_{xy}}, U_{k_{xy}} \right) \right]_{l_{xy}, k_{xy} \in R(x,y)} \tag{8}$$

3.3 Expression of the boundaries of probabilistic port hinterland

Let $\mathbf{U}_{xy}$ be the row vector of the random transportation costs of all routes in $R(x,y)$, namely,

$$\mathbf{U}_{xy} = \left(U_{1_{xy}}, U_{2_{xy}}, \cdots, U_{I_{xy}} \right) \tag{9}$$

where I_{xy} is the cardinality of $R(x,y)$. The expected value of vector $\mathbf{U}_{xy}$ is denoted by row vector $\mathbf{V}_{xy}$, where,

$$\mathbf{V}_{xy} = \left(V_{1_{xy}}, V_{2_{xy}}, \cdots, V_{I_{xy}} \right). \tag{10}$$

According to the behavior assumption, shippers around (x,y) will choose the route with the minimum cost from $R(x,y)$ to transport containers from (x,y) to *S*. According to the discrete choice theory (Daganzo, 1979; Ben-Akiva and Lerman 1985), the probability of shippers choosing port *P* can be computed by,

$$F(x,y) = \sum_{l_{xy} \in R_P(x,y)} \Pr\left[U_{l_{xy}} \le \min\left(U_{k_{xy}}, k_{xy} \in R(x,y), k_{xy} \ne l_{xy} \right) \right] \tag{11}$$

where $R_P(x,y)$ is the set of routes that traverses through P. Given that $\mathbf{U}_{xy}$ is a multivariate normally distributed random vector, the probability of shippers selecting route $l_{xy} \in R_P(x,y)$ can be computed by,

$$\Pr\left[U_{l_{xy}} \le \min_{\substack{k_{xy} \in R(x,y) \\ k_{xy} \ne l_{xy}}} \left(U_{k_{xy}} \right) \right] = \int_{u_{l_{xy}} < u_{1_{xy}}} \cdots \int_{u_{l_{xy}} = -\infty}^{u_{l_{xy}} = \infty} \cdots \int_{u_{l_{xy}} < u_{I_{xy}}} \left[(2\pi)^{I_{xy}} \left| \Sigma_{xy} \right| \right]^{-0.5} \cdot \exp\left\{ -\frac{1}{2} \left(\mathbf{u}_{xy} - \mathbf{V}_{xy} \right)' \left(\Sigma_{xy} \right)^{-1} \left(\mathbf{u}_{xy} - \mathbf{V}_{xy} \right) \right\} du_{1_{xy}} du_{2_{xy}} \ldots du_{I_{xy}} \tag{12}$$

Let parameter $\alpha \in [0,1]$ denote the probability that shippers around location (x,y) select port *P* to transport their containers to destination *S*, the probabilistic hinterland of *P* can be defined to be an area, over which shippers select *P* to transport containers with probability equal to or greater than α. This area can be mathematically formulated by set,

$$\psi_P(\alpha) = \left\{ (x,y) \mid F(x,y) \ge \alpha \right\}. \tag{13}$$

The boundary of the probabilistic hinterland of *P* corresponding to α is identified by the curve defined as

$$\bar{\psi}_P(\alpha) = \left\{ (x,y) \mid F(x,y) = \alpha \right\} \tag{14}$$

Fig. 3 shows the concept of probabilistic port hinterland and its boundary with respect to a certain probability value α. In the figure, the gray area is the hinterland of the specific port *P* and the boundary $\bar{\psi}_P(\alpha)$ is indicated by the bold curve. Within hinterland $\psi_P(\alpha)$, shippers around each point have a collection of candidate intermodal routes to deliver containers to destination *S*. Summing the probabilities of the shippers choosing the routes passing through *P* gives total probability equal to or greater than α. At each point of boundary $\bar{\psi}_P(\alpha)$, the probability that shippers choose the routes passing through *P* remains at a same value, α. It has to be noted that the piecewise linear characteristic of intermodal routes may result in a non-continuous boundary curve in the *x*-*y* plane. A method that is applicable to graphically depict boundary curves is needed.

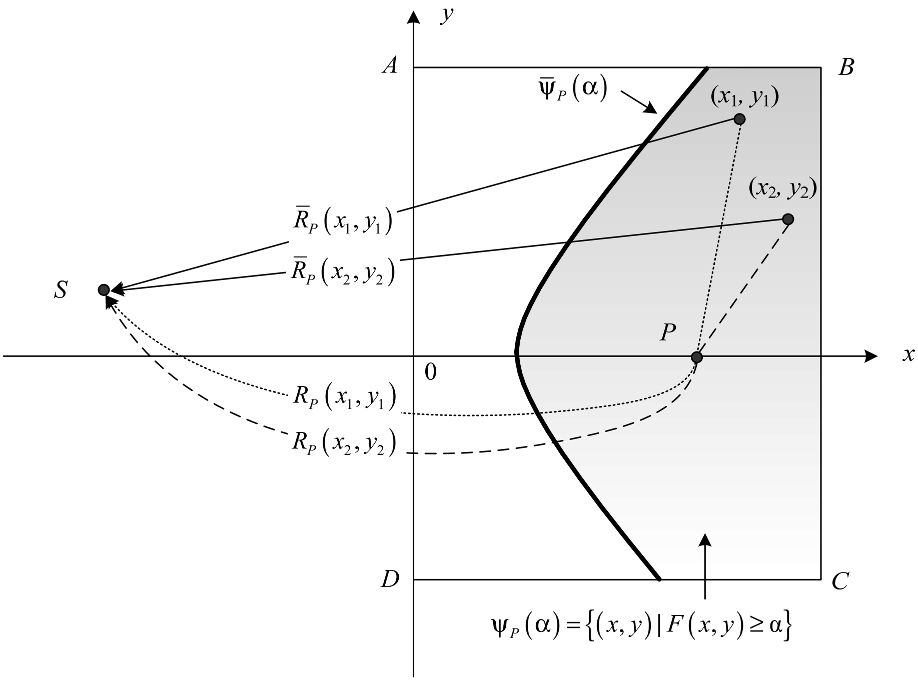


Figure 3 Illustration of the probabilistic hinterland of an interested port *P*

4 AN ALGORITHM TO ESTIMATE HINTERLAND BOUNDARIES

Let Ω be the study area containing the hinterland of port *P* in the *x*-*y* plane. To draw the hinterland boundaries, it is necessary to compute the probability of shippers choosing an intermodal route that connects origin (x, y) to destination *S* via *P*. We first discretize area Ω by setting regularly spaced horizontal and vertical lines in the *x*-*y* plane. For each point intersected by a horizontal and vertical line, the Monte Carlo simulation is employed to estimate the route choice probability of shippers. Then, the boundary curve defined by eqn. (14) can be approximated by using a curve fitting procedure. The algorithm includes four steps in the following section.

4.1 The Monte Carlo Simulation Based Algorithm

Step 1 (Discretization) Discretize area Ω by generating horizontal and vertical lines with regular space Δy between two consecutive horizontal lines and regular space Δx between two successive vertical lines. Let $\bar{\Omega}$ denote the set of all the points intersected by these horizontal lines and vertical lines.

Step 2 (Monte Carlo simulation) For each point (*x*, *y*) intersected by a horizontal line and a vertical line, i.e. $(x,y)\in\bar{\Omega}$, perform the following operations:

Step 2.1 (Mean and covariance calculation) Compute $\mathbf{V}_{xy}$ and covariance matrix $\boldsymbol{\Sigma}_{xy}$ defined by eqns. (5) and (7), respectively.

Step 2.2 (Sampling) Generate *N* pseudorandom samples of vector $\mathbf{U}_{xy}$ following multivariate normal distribution with mean value $\mathbf{V}_{xy}$ and covariance matrix $\boldsymbol{\Sigma}_{xy}$, Let $\Gamma(x,y)$ be the set of these samples, namely,

$$\Gamma(x,y)=\left\{\left(u_{1_{xy}}^{(i)},u_{2_{xy}}^{(i)},\cdots,u_{l_{xy}}^{(i)}\right)\middle|\, i=1,2,\cdots,N\right\} \tag{15}$$

Step 2.3 (Probability estimation) Estimate the probability $F(x,y)$ by estimator,

$$\hat{F}(x,y)=\sum_{l_{xy}\in R_P(x,y)}\frac{K_{l_{xy}}}{N} \tag{16}$$

In eqn.(16), $K_{l_{xy}}$ is the number of the vectors (samples) in set $\Gamma(x,y)$, in each of which the observation representing the transportation cost of route l_{xy} is the minimum among all the routes in *R*(x,y). $K_{l_{xy}}$ is the cardinality of the following subset of set $\Gamma(x,y)$:

$$\Theta(l_{xy})=\left\{\left(u_{1_{xy}},u_{2_{xy}},\cdots,u_{l_{xy}}\right)\in\Gamma(x,y)\,\middle|\, u_{l_{xy}}\le u_{k_{xy}},k_{xy}\in R(x,y),\,k_{xy}\ne l_{xy}\right\} \tag{17}$$

Step 3 (Determining the points on the boundary curve) For given probability α and a tolerable error ε_1, find a set of points denoted by $\hat{\psi}_P(\alpha)$,

$$\hat{\psi}_P(\alpha)=\left\{(x,y)\in\bar{\Omega}\,\middle|\,\alpha-\varepsilon_1\le\hat{F}(x,y)\le\alpha+\varepsilon_1\right\} \tag{18}$$

Step 4 (Boundary curve fitting) For $\hat{\psi}_P(\alpha)$ obtained in Step 3, execute the two substeps:

Step 4.1 (Cluster analysis) Partition all the points in $\hat{\psi}_P(\alpha)$ into *k* exclusive sub-sets, $\hat{\psi}_P^1(\alpha),\hat{\psi}_P^2(\alpha),...,\hat{\psi}_P^k(\alpha)$, by the *k*-means cluster analysis method. The predetermined number of clusters *k* can be identified as a priori or by the elbow criterion.

Step 4.2 (*Polynomial curve fitting*) Find a polynomial function $\hat{f}_\alpha^j(y)=\sum_{i=0}^{n}a_iy^i$ that can well fit points in the sub-set $\hat{\psi}_P^j(\alpha), j=1,2,...,k$ by the least-square estimation method.

The Monte Carlo simulation and boundary cure fitting procedure are two crucial steps in identifying the boundaries of the probabilistic port hinterland. It has been well known that the precision of probability estimation is very much related to the sample size adopted in the simulation. It would be interesting to know the computational effort needed to achieve a certain level of significance. For such a purpose, a lower bound on the sample size used in Step 2.2 is examined in the following Section 4.2. In addition, the piecewise liner characteristic of intermodal routes may result in discontinuous boundary curves. To graphically show the boundary curve, we first use cluster analysis to segment all the points in accordance with the same probability into several groups. Then, we apply the polynomial function fitting method to each of the groups to obtain boundary curves.

4.2 A Lower Bound on the Sample Size in Simulation

In Step 2.2, the sampling process is executed to generate multivariate normal samples with the given mean value and covariance matrix for each location $(x,y)\in\bar{\Omega}$. The random variable sampling can be done by using many software packages such as Matlab and Minitab. We use Matlab to generate samples in this paper.

By Step 2.3, the probability of shippers choosing a particular route $l_{xy}\in R(x,y)$, $\hat{p}_{l_{xy}}$, can be estimated by,

$$\hat{p}_{l_{xy}}=\frac{K_{l_{xy}}}{N},\forall l_{xy}\in R(x,y) \tag{19}$$

where $K_{1_{xy}},K_{2_{xy}},...,K_{I_{xy}}$ follow a multinomial distribution with *N* independent Bernoulli trials and the success probabilities $p_{1_{xy}}$, $p_{2_{xy}}$, …, $p_{I_{xy}}$. Parameter $p_{l_{xy}}$ is the probability that route l_{xy} is the minimum one among all the routes in $R(x,y)$ in terms of transportation cost. The expected value and variance of estimator $\hat{F}(x,y)$ defined in eqn.(16) can thus be calculated by,

$$\mathrm{E}\left[\hat{F}(x,y)\right]=\sum_{l_{xy}\in R_P(x,y)}\mathrm{E}\left(\hat{p}_{l_{xy}}\right)=\sum_{l_{xy}\in R_P(x,y)}p_{l_{xy}} \tag{20}$$

$$\begin{aligned}\mathrm{var}\left[\hat{F}(x,y)\right]&=\sum_{l_{xy}\in R_P(x,y)}\mathrm{var}\left(\hat{p}_{l_{xy}}\right)+\sum_{\substack{l_{xy},k_{xy}\in R_P(x,y);\\ l_{xy}\neq k_{xy}}}\mathrm{cov}\left(\hat{p}_{l_{xy}},\hat{p}_{l_{xy}}\right)\\&=\sum_{l_{xy}\in R_P(x,y)}\left[p_{l_{xy}}\left(1-p_{l_{xy}}\right)/N\right]-\sum_{\substack{l_{xy},k_{xy}\in R_P(x,y);\\ l_{xy}\neq k_{xy}}}\left(p_{l_{xy}}p_{k_{xy}}/N\right)\end{aligned} \tag{21}$$

As $\hat{F}(x,y)$ can be approximated by a normal distribution with a large *N*, the confidence interval at the significance level of 95% for estimating $F(x,y)$ can be represented by,

$$\hat{F}(x,y)\pm 1.96\sqrt{\sum_{l_{xy}\in R_P(x,y)}\left[p_{l_{xy}}\left(1-p_{l_{xy}}\right)/N\right]-\sum_{l_{xy},k_{xy}\in R_P(x,y);l_{xy}\neq k_{xy}}\left(p_{l_{xy}}p_{k_{xy}}/N\right)} \tag{22}$$

In Appendix, we show that the square root shown in eqn. (22) has an upper bound as follows:

$$SR_{\max}=\begin{cases}0.5\sqrt{1/N}, & \left|R_P(x,y)\right|=1\\ \dfrac{\sqrt{\left|R_P(x,y)\right|/N}}{\left(\left|R_P(x,y)\right|+1\right)}, & \left|R_P(x,y)\right|\geq 2\end{cases} \tag{23}$$

where $\left|R_P(x,y)\right|$ denotes the number of the routes passing though port *P* from origin (x,y) to destination *S*. Hence, we can expect (with 95% confidence) to obtain an estimation of $F(x,y)$ in Step 2.3 with an error not exceeding $\gamma\in(0,1)$ if we take the sample size:

$$N \geq \begin{cases} (1.96)^2 \times (0.5)/\gamma^2, & \left|R_P(x,y)\right| = 1 \\ (1.96)^2 \times \dfrac{\left|R_P(x,y)\right|}{\gamma^2\left(\left|R_P(x,y)\right|+1\right)^2}, & \left|R_P(x,y)\right| \geq 2 \end{cases} \tag{24}$$

For example, if $\gamma = 0.01$ and $\left|R_P(x,y)\right| = 4$, the sample size $N \geq 6146$ according to eqn. (24). To gain a higher accuracy, a larger sample size is needed.

4.3 Cluster Analysis and Polynomial Function Fitting

To deal with discontinuous boundary curves, we employ cluster analysis to segment the points in set $\hat{\psi}_P(\alpha)$ into several clusters (groups). Given the predetermined number of clusters, *k*, all the points in $\hat{\psi}_P(\alpha)$ can be partitioned into *k* exclusive groups, $\hat{\psi}_P^1(\alpha)$, $\hat{\psi}_P^2(\alpha)$, $\cdots$, $\hat{\psi}_P^k(\alpha)$, by the *k*-means cluster analysis method (MacQueen 1967) . The method minimizes the total within-cluster variance:

$$\sum_{j=1}^{k} \sum_{(x,y)\in\hat{\psi}_P^j(\alpha)} \left[\left(x-\bar{x}_j\right)^2 + \left(y-\bar{y}_j\right)^2\right] \tag{25}$$

where $\left(\bar{x}_j, \bar{y}_j\right)$ is the centroid of cluster $j\,(j=1,2,...,k)$ in the *x-y* plane.

The number of clusters, *k*, plays an important role in performing the *k*-means clustering analysis method. It can be determined as a priori according to the dispersion style of the scattering simulation points in $\hat{\psi}_P(\alpha)$ or by the elbow criterion (Ketchen and Shook 1996). Note that the elbow criterion is utilized to select an integer *k* such that adding additional clusters does not add sufficient between-cluster variance. In general, an adequate *k* can be obtained by plotting the percentage of variance explained by clusters against the number of clusters. The percentage of variance explained by clusters is defined as the proportion of the between-cluster variance to the total variance. The adequate *k* can be visually identified where the marginal gain of the percentage of variance explained by clusters abruptly falls.

After getting sets $\hat{\psi}_P^j(\alpha), j=1,2,\cdots,k$, we could use the classical least-square method to find polynomial functions that well fit the points in the sets. These functions are regarded as an approximation of the boundary curve corresponding to probability α. In other words, given a subset $\hat{\psi}_P^j(\alpha)$, we need to determine coefficients a_i, $i=0,1,\cdots,n$, and the highest order n of the polynomial function $f_\alpha^j(y)=\sum_{i=0}^{n} a_i y^i$ such that the following curve fitting error is minimized for each $j=1,2,...,k$.

$$Error_n^j\left(n,a_0,...,a_n\right) = \sum_{(x,y)\in\hat{\psi}_P^j(\alpha)} \left(x - \sum_{i=0}^{n} a_i y^i\right)^2 \tag{26}$$

The order *n* can be gradually increased to obtain a better fitting function with the stop criterion:

$$\left[\min\left(Error_n^j\right)-\min\left(Error_{n+1}^j\right)\right]/\min\left(Error_n^j\right)\le \varepsilon_2 \tag{27}$$

where ε_2 represents a tolerable error in the curve fitting process. Let n^* and $a_i^*, i=1,2,...,n$ be the minimizer for eqn. (26), i.e.,

$$\left(n^*,a_0^*,...,a_n^*\right)=\arg\min\left[\sum_{(x,y)\in\hat{\psi}_P^j(\alpha)}\left(x-\sum_{i=0}^{n}a_i y^i\right)^2\right] \tag{28}$$

The boundary based on $\hat{\psi}_P^j(\alpha)$ can be represented by the fitting function:

$$f_\alpha^j(y)=\sum_{i=0}^{n^*}a_i^* y^i \,. \tag{29}$$

5 ILLUSTRATIVE EXAMPLES

The proposed modeling approach provides a useful tool to estimate the boundaries of the probabilistic hinterland of a port for both policymakers and port operators. In this section, two illustrative examples are given to show the applicability of the proposed approach. Fig. 4 shows the intermodal network involving three intermodal routes. Based on the network, we intend to estimate the boundaries of the hinterland of Shanghai port by considering China and Mainland Southeast Asia (CMSA) as study area and a point in Myanmar as destination *S*.

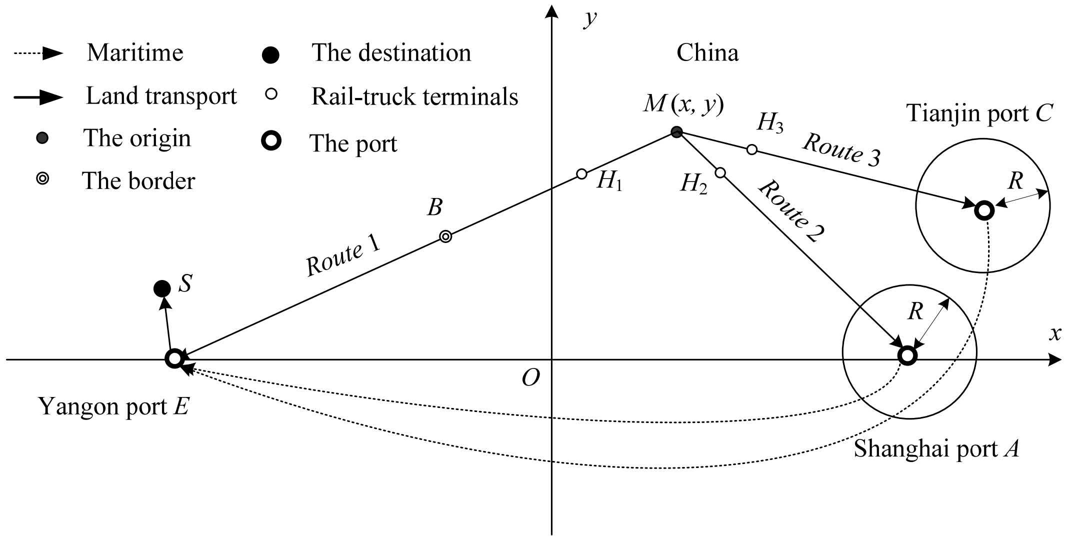


Figure 4 An intermodal freight transportation network example

Fig. 4 actually shows an intermodal network that crosses several territories in CMSA, such as, China, Myanmar, and the countries along the Strait of Malacca. Intermodal routes generally integrate multi-mode transportation services, such as short-haul truck services, transshipments at ports and maritime transportation. For instance, the delivery of containers from China to Myanmar can be achieved by the following procedure: (i) containers are first transported to a home port of China by using truck services or a combination of truck and rail services, (ii) and then to the port of Yangon in Myanmar by maritime mode via the Strait of Malacca. Alternatively, shippers can also choose the China-Myanmar landbridge (Route 1) that integrates short-haul truck and long-haul rail services to transport their containers.

In Fig. 4, nodes H_1, H_2 and H_3 represent rail-truck terminals. Node *B* denotes a border crossing terminal between China and Myanmar. Shanghai port, Tianjin port and Yangon port are denoted by nodes *A* = (2171km, 0), *C* = (2171 km, 950 km) and *E* = (-2171 km, 0), respectively. The study area can be defined by,

$$\Omega = \{(x,y) | x \in [0, 3{,}000 \text{ km}], y \in [-1{,}500 \text{ km}, 1{,}500 \text{ km}]\} \tag{30}$$

Containers will be transported from origin *M* = (*x*, *y*) to destination *S*, which is situated in a circular area with radius *R* = 200 km and centered at node *E*. Three intermodal routes, Routes 1, 2 and 3, are available to accomplish the delivery. Route 1 directly transports containers from node *M* to *E* by the landbridge integrating truck service MH_1, rail service H_1E, and border crossing process at node *B*. Finally, the containers are transported to *S* by truck service. Route 2 combines land transport *MA*, maritime transport *AE* and short-haul truck service *ES*. Route 3 transports containers to *S* by passing through the port of Tianjin in China. The three routes share the common link *ES*. The expected transportation costs of the three routes are calibrated based on historical data, as discussed in the following Section 5.1.

5.1 Transportation Cost Calibration

If origin *M* is more than 200 km away from *A* or *C*, a domestic intermodal route combining truck, rail and rail-truck transfer in China is utilized to transport containers to *A* or *C*; only truck service is used otherwise. Available data obtained based on North American landbridge operations (Shipmentlink 2008) are used for calibrating the transportation cost of landbridge transprot in CMSA.

5.1.1 Costs of short-haul truck and long-haul rail services

Trucking transport is a dominant solution for short-haul cargo delivery due to its ability to provide door-to-door services. It is estimated that delivering one container load unit (TEU or FEU) for 100 km by truck takes about 150 USD in the USA (USCC 2006). The charge for transporting about 200 km can thus be estimated at 300 USD. We assume that the cost for transporting one TEU for a distance not exceeding 200 km remains 300 USD. We obtain the following cost-distance function for long-haul rail transport from Wang et al. (2009).

$$c(l) = 268 + 0.267l \quad R^2 = 0.717 \tag{31}$$

where *l* denotes travel distance and $c(l)$ is the expected transportation cost corresponding to *l* .

5.1.2 Costs of rail-truck and border crossing operations

The constant value 268 USD in eqn. (31) is interpreted as the charge incurred for inventory, loading and unloading operations at a rail-truck terminal. UNESCAP (2003) shows that the cost incurred in the border crossing process between China and Mongolia is about 293 USD. The border crossing time is around one to five days. We thus take 293 USD as the border crossing cost between China and Myanmar.

5.1.3 Costs of maritime transport and port operations

The costs of maritime transport and container handling processes at Shanghai port *A*, Tianjin port *C* and Yangon port *E* can be estimated by using the data from OOCL (2008), ASEAN (2001), UNCTD (2007), Shipmentlink (2008) and MPA (1998). These data are tabulated in the second and fourth column of Table 1. In Table 1, the symbol *C* with a subscripted letter representing the name of corresponding component denotes transportation cost incurred on the component. For instance, C_{H_1E} denotes the cost incurred on link H_1E.

Table 1 Expected values and variances of three intermodal routes

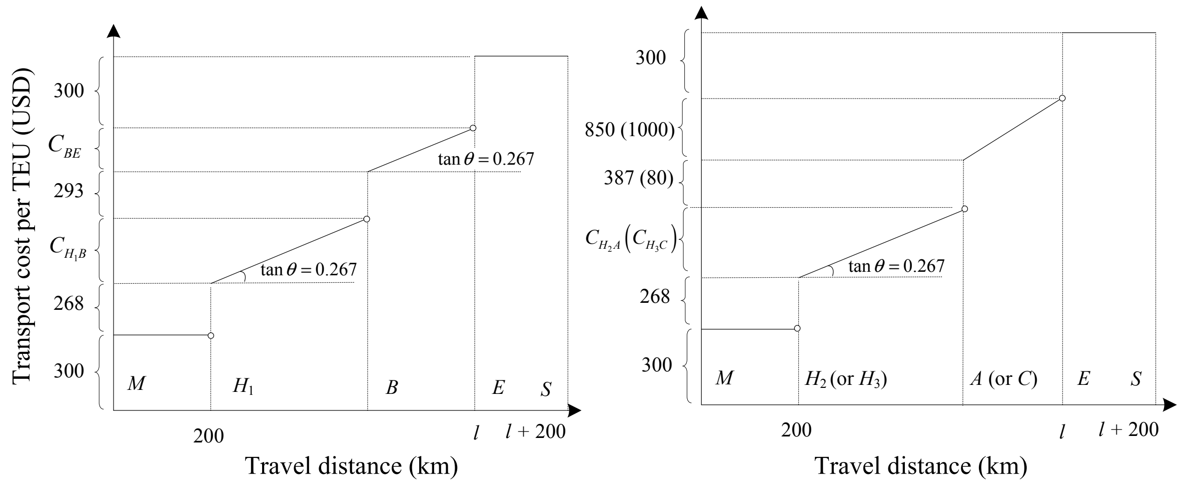
Route 1	**Cost (USD)**	σ_C^2	**Route 2 (Route 3)**	**Cost (USD)**	σ_C^2
MH_1	300	60.0^2	MH_2 (MH_3)	300	60.0^2
At H_1	268	53.6^2	At H_2 (H_3)	268	53.6^2
H_1E	C_{H_1E}	$\left(0.2C_{H_1E}\right)^2$	H_2A (H_3C)	$C_{H_2A}\left(C_{H_3C}\right)$	$\left(0.2C_{H_2A}\right)^2\left[\left(0.2C_{H_3C}\right)^2\right]$
At *B*	293	58.6^2	At *A* (*C*)	387 (80)	77.4^2 (16^2)
ES	300	60.0^2	*AE* (*CE*)	850 (1000)	170^2 (200^2)
			ES	300	60.0^2

Note: when the port *A* or *C* is less than 200 km away from *M*, only the cost for short-haul truck service is involved in the domestic land transport

5.1.4 Piecewise linear cost functions of the three routes

As shown in Fig. 5, the piecewise linear functions for the expected transportation costs of the three intermodal routes can be determined based on the data shown in Table 1 and long-haul rail transport cost expressed by eqn. (31).

In Fig. 5, each of the piecewise linear curves comprises several components representing the costs incurred at nodes and links. The abrupt changes in a curve represent the costs expended at ports, borders or rail-truck terminals. The linear parts in the curves represent the costs incurred on links. Due to the loss of available and good data, the standard deviation of the cost incurred at a node or on a link is estimated at 20% of its mean value. The covariance of the three intermodal routes in terms of cost is determined by the variance of their common link *ES*. Table 1 shows the expected cost of each component of the three intermodal routes and its variance.



(*a*): cost for the intermodal route 1 (*b*): cost for the intermodal route 2 (or route 3)

Figure 5 The piecewise linear functions for three intermodal routes

5.2 Example 1- Two Competing Routes

In this example, only Routes 1 and 2 are considered for boundary estimation. As the transportation costs of Routes 1 and 2 follow a bivariate normal distribution, we can luckily derive the boundaries of the probabilistic hinterland of Shanghai port in an analytical way. Based on eqns. (12)-(13), within the circular area with 200 km as radius and centered at Shanghai port *A*, the boundary with respect to α can be represented by,

$$\psi_P(\alpha) = \left\{ (x,y) \left| \frac{C_{H_1E} - 676}{\sqrt{48400 + \left(0.2C_{H_1E}\right)^2}} = \Phi^{-1}(\alpha) \right. \right\} \tag{32}$$

where $\Phi^{-1}(\cdot)$ is the reverse cumulative distribution function of the standard normal distribution, and

$$C_{H_1E} = 0.267\left(\sqrt{(x+2171)^2 + y^2} - 200\right) \tag{33}$$

Outside the circle, the boundary can be expressed by,

$$\psi_P(\alpha) = \left\{ (x,y) \left| \frac{C_{H_1E} - C_{H_2A} - 944}{\sqrt{51271 + \left(0.2C_{H_1E}\right)^2 + \left(0.2C_{H_2A}\right)^2}} = \Phi^{-1}(\alpha) \right. \right\} \tag{34}$$

where ,

$$C_{H_2A} = 0.267\left(\sqrt{(x-2171)^2 + y^2} - 200\right) \tag{35}$$

We are also interested in approximating the boundaries by using the simulation algorithm, and comparing the simulation result with the analytical one. To run the simulation, the study area is divided into two parts by the circle with radius *R* = 200 km and centered at *A*. Inside the circle, the expected transportation costs of Routes 1 and 2 can be expressed by,

$$\mathbf{V}_{xy} = \left(V_{1_{xy}} = C_{H_1E} + 1161, V_{2_{xy}} = 1837\right)' \tag{36}$$

The covariance matrix for the two routes is given by,

$$\Sigma_{xy} = \begin{bmatrix} 13507 + \left(0.2C_{H_1E}\right)^2 & 3600 \\ 3600 & 42093 \end{bmatrix} \tag{37}$$

Outside the circle, the expected transportation costs and covariance matrix of the two routes are represented as follows,

$$\mathbf{V}_{xy} = \left(V_{1_{xy}} = C_{H_1E} + 1161, V_{2_{xy}} = C_{H_2A} + 2105\right)' \tag{38}$$

$$\Sigma_{xy} = \begin{bmatrix} 13507 + \left(0.2C_{H_1E}\right)^2 & 3600 \\ 3600 & 44964 + \left(0.2C_{H_2A}\right)^2 \end{bmatrix} \tag{39}$$

The algorithm is coded using MATLAB and is executed by using a desktop with CPU of Pentium 4 3.00 GHz and 4G RAM. Let parameters $\Delta x = \Delta y = 5$ km, $\varepsilon_1 = 10^{-4}$, $\varepsilon_2 = 10^{-8}$, $\gamma = 0.01$ and *N* = 20,000. We adopt *k* = 1 for the cluster analysis in the solution algorithm in order to obtain the polynomial fitting curves with respect to α = 0.1, 0.2, 0.5 and 0.90. Fig. 6 shows the resultant fitting curves by the simulation algorithm as well as the curves representing the analytical expressions defined in eqns. (32) and (34).

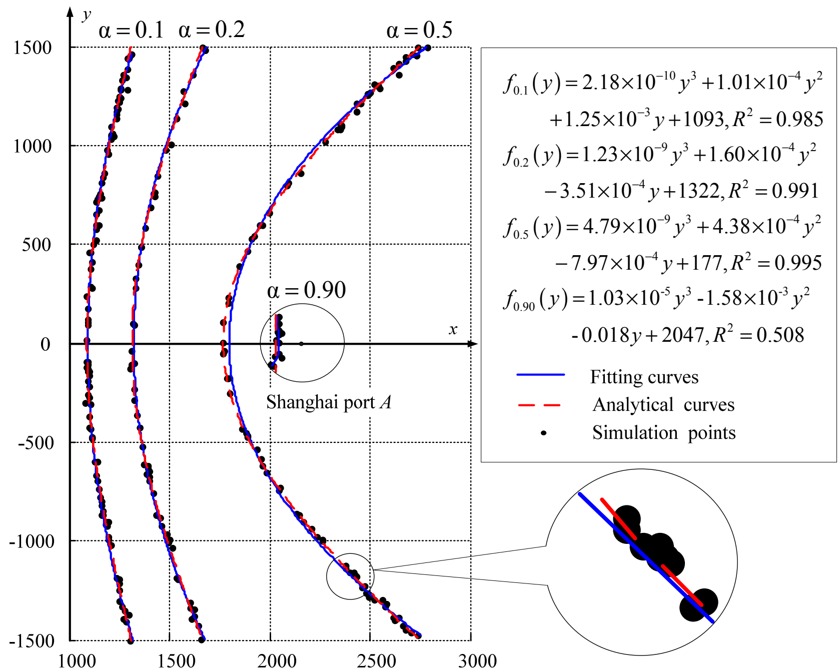


Figure 6 The analytical and numerical hinterland boundary curves in Example 1

It can be seen that the fitting curves obtained by the simulation algorithm are nearly coincident with the curves obtained by the analytical closed-form expressions. The coincidence demonstrates the applicability of the proposed solution algorithm to some extent.

5.3 Example 2- Three Competing Routes

We now incorporate another Route 3 to see what will happen to the boundaries of the hinterland of Shanghai port. The study area is now partitioned into three subareas by two circles: one has radius of 200 km and is centered at Shanghai port *A* and the other has the same radius but is centered at Tianjin port *C*.

If origin *M* = (*x*, *y*) falls in circular area around *A*, the expected values and covariance matrix of the three routes are represented by,

$$\mathbf{V}_{xy} = \left(V_{1_{xy}} = C_{H_1E} + 1161, V_{2_{xy}} = 1837, V_{3_{xy}} = C_{H_3C} + 1948\right)' \tag{40}$$

$$\Sigma_{xy} = \begin{bmatrix} 13507 + \left(0.2C_{H_1E}\right)^2 & 3600 & 3600 \\ 3600 & 42093 & 3600 \\ 3600 & 3600 & 50328 + \left(0.2C_{H_3C}\right)^2 \end{bmatrix} \tag{41}$$

where C_{H_1E} is given by eqn. (33) and C_{H_3C} is expressed by

$$C_{H_3C} = 0.267\left[\sqrt{\left(x-2171\right)^2 + \left(y-950\right)^2} - 200\right]. \tag{42}$$

If origin (*x*, *y*) is located in the circular area centered at *B*, we have

$$\mathbf{V}_{xy} = \left(V_{1_{xy}} = C_{H_1E} + 1161, V_{2_{xy}} = C_{H_2A} + 2105, V_{3_{xy}} = 1680\right)' \tag{43}$$

$$\Sigma_{xy} = \begin{bmatrix} 13507 + \left(0.2C_{H_1E}\right)^2 & 3600 & 3600 \\ 3600 & 44964 + \left(0.2C_{H_2A}\right)^2 & 3600 \\ 3600 & 3600 & 47456 \end{bmatrix} \tag{44}$$

If origin (*x*, *y*) is in the area outside the two circles, we have,

$$\mathbf{V}_{xy} = \left(V_{1_{xy}} = C_{H_1E} + 1161, V_{2_{xy}} = C_{H_2A} + 2105, V_{3_{xy}} = C_{H_3C} + 1948\right)' \tag{45}$$

$$\Sigma_{xy} = \begin{bmatrix} 13507 + \left(0.2C_{H_1E}\right)^2 & 3600 & 3600 \\ 3600 & 44964 + \left(0.2C_{H_2A}\right)^2 & 3600 \\ 3600 & 3600 & 50328 + \left(0.2C_{H_3C}\right)^2 \end{bmatrix} \tag{46}$$

Set parameters $\Delta x = \Delta y = 5$ km, $\varepsilon_1 = 10^{-4}$, $\varepsilon_2 = 10^{-8}$, $\gamma = 0.01$ and *N* = 30000. For probabilities α = 0.015, 0.1, 0.2 and 0.80, we let *k* = 1 and obtain four fitting curves representing hinterland boundaries corresponding to these probabilities as shown in Fig. 7.

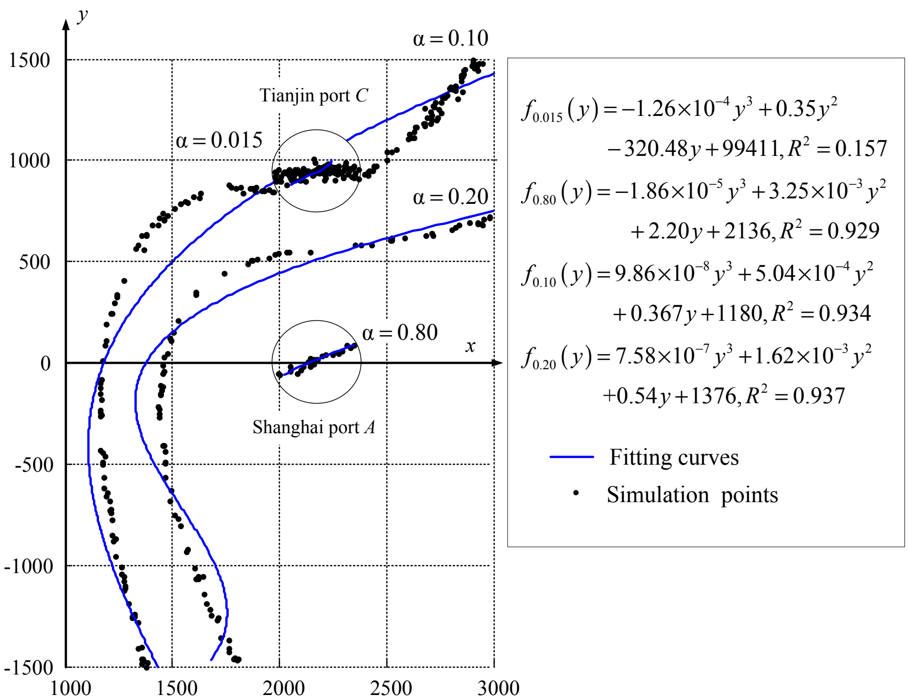
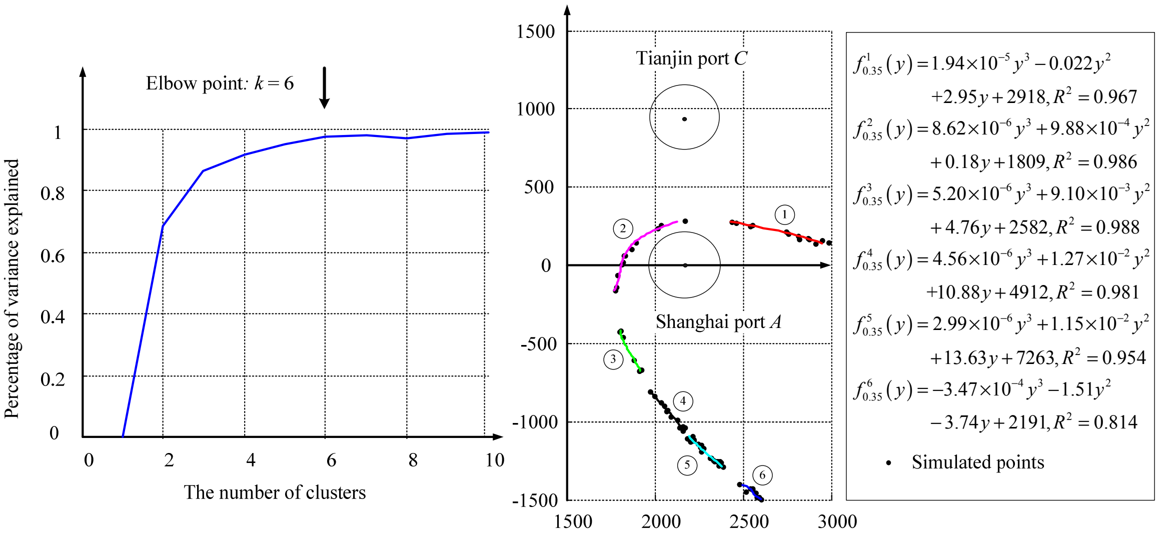


Fig. 7. The hinterland boundary curves in Example 2

Fig. 8a shows the elbow criterion obtained in cluster analysis for α = 0.35. According to Fig. 8a, *k* = 6 should be adopted, which implies that the points in $\hat{\psi}_P(\alpha = 0.35)$ should be classified into six groups. Fig. 8b shows six polynomial curves that are obtained by fitting the points in set $\hat{\psi}_P(\alpha = 0.35)$. These curves represent the boundaries of the probabilistic hinterland of Shanghai port when $\alpha = 0.35$. The figure demonstrates the applicability of the developed approach when discontinuous hinterland boundaries exist.



(*a*) Elbow criterion for the number of clusters (*b*) Fitting curves corresponding to probability 0.35

Fig. 8. Elbow criterion and curving fitting for the probability 0.35

5.4 Findings and Discussions

The fitting curves as shown in Figures 6, 7 and 8 represent the boundaries of probabilistic hinterland of Shanghai port. These curves actually construct a series of contours to specify the spatial area, over which shippers choose Shanghai port to handle their containers with certain probabilities. With these contours, port operators or other related decision makers are able to make well-informed policies to enhance port competitiveness.

In addition, the significance of travel distance to the hinterland of a port can be envisioned from the figures. The travel distance to a specific port is negatively related to the probability of shippers choosing the port. In particular, since truck service is the only mode utilized for short-haul transportation inside the circular area centered at the port and it is generally economical and efficient, most of shippers in the area will use truck services to transport containers to Shanghai port before going to maritime mode. This result also indicates the importance of an efficient and economical short-haul truck service in expanding the extent of a port's hinterland. Another easy observation tells that, with the incorporation of Route 3, the hinterland of Shanghai port corresponding to same probability will decrease. This result is intuitive and further justifies the applicability of the proposed approach.

6. CONCLUSIONS

In this paper we proposed a new definition of probabilistic port hinterland for intermodal freight transportation operation systems. After investigating the piecewise characteristic of intermodal routes, we developed mathematical expressions of the probabilistic hinterland of a specific port and its boundaries with respect to certain probabilities. To graphically show the hinterland boundaries, a Monte Carlo simulation based algorithm, which integrates a cluster analysis method and boundary curve fitting method, was proposed. Moreover, we derived a lower bound on the sample size that is needed in the simulation algorithm in order to achieve a certain level of significance. Two illustrative examples were finally given to demonstrate the applicability of the proposed model and algorithm.

ACKNOWLEDGMENTS

This study is supported by the research grant WBS Nos. R-302-000-014-720 and R-702-000-007-720 from the NOL (Neptune Orient Lines) Fellowship Programme of Singapore.

APPENDIX. AN UPPER BOUND FOR THE SQUARE ROOT

$$SR = \sqrt{\sum_{l_{xy} \in R_P(x,y)} \left[p_{l_{xy}} \left(1 - p_{l_{xy}}\right) \Big/ N \right] - \sum_{\forall l_{xy}, k_{xy} \in R_P(x,y); l_{xy} \neq k_{xy}} \left(p_{l_{xy}} p_{k_{xy}} / N \right)}$$

Let a row vector $\mathbf{P}_{l_{xy}} = \left(p_{l_{xy}} : l_{xy} \in R_P(x,y) \right)$ and a multivariate function:

$$Z\left(\mathbf{P}_{l_{xy}}\right) = \sum_{l_{xy} \in R_P(x,y)} p_{l_{xy}} \left(1 - p_{l_{xy}}\right) - \sum_{\forall l_{xy}, k_{xy} \in R_P(x,y); l_{xy} \neq k_{xy}} p_{l_{xy}} p_{k_{xy}} \tag{A.1}$$

We build a linearly constrained strictly concave maximization problem as follows:

$$\max Z\left(\mathbf{P}_{l_{xy}}\right) = \sum_{l_{xy} \in R_P(x,y)} p_{l_{xy}} \left(1 - p_{l_{xy}}\right) - \sum_{\forall l_{xy}, k_{xy} \in R_P(x,y); l_{xy} \neq k_{xy}} p_{l_{xy}} p_{k_{xy}} \tag{A.2}$$

subject to

$$\sum\nolimits_{l_{xy} \in R_P(x,y)} p_{l_{xy}} \leq 1 \tag{A.3}$$

$$p_{l_{xy}} \geq 0, \forall l_{xy} \in R_P(x,y) \tag{A.4}$$

In the case of $\left|R_P(x,y)\right| = 1$, where $\left|R_P(x,y)\right|$ denotes the cardinality of set $R_P(x,y)$, it is straightforward to check that the minimization model (A.2) – (A.4) has the optimal objective function value:

$$Z_{\max} = 1/4 \tag{A.5}$$

If $\left|R_P(x,y)\right| \geq 2$, the optimal solution of the concave maximization model (A.2) – (A.4) should fulfill the Karush-Kuhn-Tucker (KKT) conditions (Bazaraa et al. 2006):

$$2p_{l_{xy}} - 1 + \sum\nolimits_{k_{xy} \in R_P(x,y), k_{xy} \neq l_{xy}} p_{k_{xy}} + \beta - \mu_{l_{xy}} = 0, \quad \forall l_{xy} \in R_P(x,y) \tag{A.6}$$

$$p_{l_{xy}} \mu_{l_{xy}} = 0, \quad \forall l_{xy} \in R_P(x,y) \tag{A.7}$$

$$\beta\left(\sum\nolimits_{l_{xy} \in R_P(x,y)} p_{l_{xy}} - 1\right) = 0 \tag{A.8}$$

$$\mu_{l_{xy}} \geq 0, \quad \forall l_{xy} \in R_P(x,y) \tag{A.9}$$

$$\beta \geq 0 \tag{A.10}$$

$$\sum\nolimits_{l_{xy} \in R_P(x,y)} p_{l_{xy}} \leq 1 \tag{A.11}$$

$$p_{l_{xy}} \geq 0, \quad \forall l_{xy} \in R_P(x,y) \tag{A.12}$$

Solving eqns. (A.6)-(A.12) yields the optimal solution and Lagrangian multipliers as follows:

$$p^*_{l_{xy}} = 1 / \left(\left|R_P(x,y)\right| + 1\right), \quad \forall l_{xy} \in R_P(x,y) \tag{A.13}$$

$$\mu^*_{l_{xy}} = 0, \quad \forall l_{xy} \in R_P(x,y) \tag{A.14}$$

$$\beta^* = 0 \tag{A.15}$$

Substituting the optimal solution shown in eqn. (A.13) into the objective function of the maximization model, it follows that

$$Z_{\max} = \left|R_P(x,y)\right| / \left(\left|R_P(x,y)\right| + 1\right)^2 \tag{A.16}$$

According to eqns. (A.5) and (A.16), it follows that

$$SR_{\max} = \begin{cases} 0.5\sqrt{1/N}, & \left|R_P(x,y)\right| = 1 \\ \dfrac{\sqrt{\left|R_P(x,y)\right|/N}}{\left(\left|R_P(x,y)\right|+1\right)}, & \left|R_P(x,y)\right| \geq 2 \end{cases} \tag{A.17}$$

REFERENCES

Association of Southeast Asian Nations (ASEAN) (2001) Facts and figures: cost of investing and doing business in ASEAN. ASEAN, Jakarta.

Bazaraa MS, Sherali HD, Shetty CM (2006) Nonlinear programming: theory and algorithms, 3rd edn. Hoboken, Wiley-Interscience, New Jersey.

Ben-Akiva, M. and Lerman, S. R., 1985. *Discrete choice analysis: theory and application to travel demand*. Cambridge, Massachusetts: MIT Press.

Chang YT, Lee SY and Tongzon JL (2008) Port selection factors by shipping lines: different perspectives between trunk liners and feeder service providers. Marine Policy 32(6): 877-885.

Daganzo, C., 1979. *Multinomial probit: the theory and its application to demand forecasting*. New York: Academic Press.

De Lange PW (2007) Port competition and selection in contestable hinterlands. *European Journal of Transportation and Infrastructure Research* 7(1): 1-14.

Fetter FA (1924) The economic law of market areas*. The Quarterly Journal of Economics* 38 (3): 520-529.

Garrido RA, Leva M (2004) Port of destination and carrier selection for fruit exports: a multi-dimensional space-time multi-nominal probit model. *Transportation Research Part B* 38(7): 657-667.

Hoare A (1986) British ports and their hinterland: a rapidly changing geography. *Geografiska Annaler* 68B(4): 29-40.

Hyson CD, Hyson WP (1950) The economic law of market areas. *The Quarterly Journal of Economics* 64 (2): 319-327.

Ketchen DJ, Shook CL (1996) The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal* 17: 441-458.

Lee SW, Song DW, Ducruet C (2008) A tale of Asia's word ports: the spatial evolution in global hub port cities. *Geoforum* 39: 372-385.

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. Proc of the fifth Berkeley Symp on Math Statistics and Probab. Berkeley, University of California Press 1: 281-297.

Malchow M, Kanafani A (2001) A disaggregate analysis of factors influencing a port's attractiveness. *Maritime Policy and Management* 28(3): 361-373.

Malchow M, Kanafani A (2004) A disaggregate analysis of port selection. *Transportation Research Part E* 40: 317-337.

Meng Q, Wang X (2010) Utility-Based estimation of probabilistic port hinterland for networks of intermodal freight transportation. *Trans Research Record: Journal of the Trans Research Board* 2168: 53-62.

Myanmar Port Authority (MPA) (1998) Tariff and dues. http://www.mot.gov.mm/mpa/services.html. Accessed 21 October 2012.

Niérat P, (1997) Market area of rail-truck terminals: pertinence of the spatial theory. *Transportation Research Part A* 31 (2): 109-127 .

Notteboom T, Rodrigue JP (2007) Re-assessing port-hinterland relationships in the context of global commodity chains. In Wang JJ (eds), Ports, Cities, and Global Supply Chains, Chapter 5: 51-66, Ashgate, London.

O'Kelly ME, Miller HJ (1989) A synthesis of some market area delimitation models. Growth Chang 20 (3): 14-33.

Orient Overseas Container Line Limited (OOCL) (2008) OOCL rate tariff. http:// www .oocl.com /eng /ourservices/eservices/tariffandrates/. Accessed 21 October 2012.

Parola, F. and Sciomachen, A, 2005 Intermodal container flows in a port system network: Analysis of possible growths via simulation models. *International Journal of Production Economics*, 97:75-84.

Patel JK, Read CB (1996) *Handbook of the normal distribution*. CRC Press, New York.

ShipmentLink (2008) *Tariff and rates*. http://www.shipmentlink.com/tvs2/download _txt/AUE _E.txt. Accessed 21 October 2012.

Tang LC, Low JMW, Lam SW (2011) Understanding port choice behavior – A network perspective. Networks and Spatial Economics 11(1): 65-82.

The U.S. Chamber of Commerce (USCC) (2006) Land transport options between Europe and Asia: commercial feasibility study. http://www.amcham.ge/res/various/Final_ Ramboll_ Study.pdf. Accessed 21 October 2012.

Tongzon JL (2008) Port choice and freight forwarders. *Transportation Research Part E* 45(1):186-195.

Tongzon, J., Chang, Y.-T., and Lee, S.-Y., 2009. How supply chain oriented is the port sector? *International Journal of Production Economics*, 122:21-34.

United Nations Conference on Trade and Development (UNCTD) (2007) Review of maritime transport 2007. http://www.unctad.org/Templates/Webflyer.asp?intItem ID=4398. Accessed 21 October 2012.

United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) (2003) Transit transport issues in landlocked and transit developing. http://www.unescap.org /59/e/E1282e.pdf . Accessed 21 October 2012.

United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) (2005) Free Trade Zone and Port Hinterland Development. http://www.unescap.org/publications/detail.asp?id=1140. Accessed 21 October 2011.

Van Cleef E (1941) Hinterland and umland. Geographical Review 31(2): 308-311.

Van Klink HA, Van de Berg GC (1998) Gateways and intermodalism. *Journal of Transport Geography* 6(1): 1-9.

Wang X, Meng Q, Miao L, Fwa TF (2009) Impact of landbridge on port market area: model development and scenario analysis. *Transportation Research Record: Journal of the Transportation Research Board* 2097: 78-87.

Wang X. and Meng Q. (2011) The impact of landbridge on the market shares of Asia Ports. *Transportation Research Part E* 47:190-203.

Wang, W. F. and Yun, W. Y., 2011. Scheduling for inland container truck and train transportation *International Journal of Production Economics* doi:10.1016/j.ijpe.2011.10.016.

Xie X (2009) An integrated sea-land transportation system model and its theory. *Transportation Research Part C* 17(4): 394-411.

Yeo G. T., Roe M, Dinwoodie J (2008) Evaluation the competitiveness of container ports in Korea and China. *Transportation Research Part A* 42: 910-921.

Zan Y (1999) Analysis of container port policy by the reaction of an equilibrium shipping market. *Maritime Policy Management* 26(4):369-381.