Characterizing the evolution of the Yangtze River Delta multi-port system using Compositional Data techniques

Dong Huang[a, b *], Manel Grifoll[a,b], Hongxiang Feng[c], Maribel Ortego[b], Pengjun Zheng[c]

[a] *Barcelona Innovation in Transport (BIT), Barcelona School of Nautical Studies, Universitat Politècnica de Catalunya - BarcelonaTech, 08003, Barcelona, Spain*

[b] *Department of Civil and Environmental Engineering (DECA), Universitat Politècnica de Catalunya–BarcelonaTech (UPC), 08034 Barcelona, Spain;*

[c] *Faculty of Maritime and Transportation, Ningbo University, Ningbo, 315832, China*

[*] *Correspondence: dong.huang@upc.edu; Tel.: +0034-697695187*

Characterizing the evolution of the Yangtze River Delta multi-port system using Compositional Data techniques

Abstract

The Yangtze River Delta multi-port system (YMPS) receives increasing attention due to its relevance in world trade and excellent competitiveness in the container traffic market. To insight into the development pattern of YMPS, this paper proposes a method that combines the Hierarchical Clustering with compositional data (CoDa) exploratory tools (i.e., biplot and dendrogram) to explore the temporal and spatial evolution of the YMPS from 1992 to 2019. CoDa describes parts of some whole (i.e., frequency and percentage), conveying relative information in the ratios between its components. Container traffic share in a multi-port region is typical CoDa. Traditional statistical approaches to CoDa could lead to spurious correlations and erroneous conclusions. However, using suitable CoDa techniques, such as the centered log-ratio (*clr*) transformation, can effectively avoid these misinterpretations. The novel method can find the temporal and spatial characteristics simultaneously. The findings indicate that the development of the YMPS has gone through four stages and the evolution of the YMPS is characterized by a tendency towards a "multi-core development" and faces a differentiated pattern of "peripheral port challenges". The analysis further improves the port system's evolutionary model and explains the underlying reason for the YMPS development. CoDa techniques also provides a new perspective for the temporal and spatial evolution of the transport discipline.

Keywords: The Yangtze River Delta multi-port system, Compositional Data, Concentration Indexes, Hierarchical Clustering.

1 Introduction

The Yangtze River Delta region is an influential intersection of the "21st Century Maritime Silk Road" and the "Yangtze River Economic Belt" (Cao et al. 2019). Meanwhile, the Yangtze River Delta region

is located in the middle of China's coastline and represents a significant gateway connecting mainland China and the world (see Figure 1). It receives more and more attention due to its relevance in world trade and excellent competitiveness in the container traffic market. This region includes Shanghai, Jiangsu and Zhejiang provinces, which account for 3.8% of China's area. However, 16.7% of China's population, 20% of China's GDP and 36.27% of China's total container throughput came from the Yangtze River Delta region in 2018 (Cao et al. 2019). Among them, Shanghai port (SH) and Ningbo port (NB) ranked first and third respectively in the world's top ten container ports in 2018, and the other thirteen ports (OTH) also exhibit excellent competitiveness. So, in recent years, the spatial and temporal evolution process of the YMPS receives increasing attention.

There are many methods to investigate the evolution of port traffic. For instance, Notteboom (1997) utilized concentration index (H* and Gini coefficient) to demonstrate that the containerization of the European ports would lead to further concentration. The European port system and most of its multi-port gateways were still undergoing a deconcentration process. Svindland, Monios, and Hjelle (2019) utilized concentration index and semi-structured interviews to explore the evolution of the Norway port system, which proved that the Norway port system followed the same concentration process then deconcentration as major port ranges. Grifoll, Karlis, and Ortego (2018) investigated the container traffic share evolution of the Mediterranean port using Hierarchical Clustering and concentration indexes, which provided an excellent method to explore the temporal evolution in a multi-port region. Pallis, George, and Vaggelas (2017) introduced the evolution of Greece's container port market and the reform process over the last decade. Research demonstrated that the Greek port system was different from the models endorsed in other countries. Recently, Feng et al. (2020) employed the ternary diagram with concentration ratios (CR(n)), H* and Aitchison distance to study inequality, shift volumes flows and competition in the port system, this method provides a new perspective for transport discipline. Those contributions have achieved considerable skills in multi-port traffic temporal evolution, but developing and applying robust and coherent analytical tools still deserves more attention to providing a conclusive characterization of temporal and spatial evolution in multi-port systems. This paper proposes a method to combine the Hierarchical Clustering based on the *clr*-transformation with CoDa exploratory tools (i.e., biplot and dendrogram) to explore the temporal and spatial evolution of the YMPS from 1992 to 2019. The novel method can find the temporal and spatial

characteristics simultaneously. So, in this contribution we will prove that through this method we can distinguish a differentiated pattern that other methods cannot meet obtaining coherent temporal periods and identifying differentiated temporal evolution from specific ports in YMPS. In this sense, the insight gained by data science may also contribute substantially to port management and policies (e.g. Chan et al., 2017; Parola et al., 2020).

Compositional data (CoDa) is quantitative descriptions of the parts of some whole (i.e., frequency and percentage), conveying relative information in the ratios between its components (see Eq.1) (Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2015). CoDa techniques have been applied to a wide variety of scientific disciplines, such as geography (Buccianti and Grunsky 2014), economics (Ferrer, Coenders, and Martínez-Garcia 2015), archaeometry (Baxter and Freestone 2006) and chemistry (Reimann et al. 2012), among others. CoDa techniques were applied by Grifoll, Ortego, and Egozcue (2019) to build port associations and reveal the underlying tendencies, offering a better interpretation of container traffic evolution. However, they conclude that further exploration on CoDa method in multi-port systems and application of advanced CoDa techniques (e.g. Sequential Binary Partition) is required to postulate new methods to gain insight on port systems and transformation. Container traffic share in YMPS is an excellent objective to explore the applicability of CoDa techniques due to its relevance in world trade and excellent competitiveness in the container traffic market. We believe that it is necessary to study the temporal and spatial evolution of the container traffic share in the YMPS. It is also beneficial to understand the underlying impact of policies on the development of the port.

The contributions of this paper are three folds. Firstly, a novel method is proposed to combine the Hierarchical Clustering based on the *clr*-transformation with CoDa exploratory tools (i.e., biplot and dendrogram) to investigate the temporal and spatial evolution of the YMPS from 1992 to 2019. Unlike the traditional concentration index (i.e., H*, CR(n) and Gini coefficient), the novel method can simultaneously explore temporal and spatial characteristics and find the differentiated development pattern that other methods cannot meet. In this sense, this method contributes further to improve the port system's evolutionary model and provides a new perspective for the temporal and spatial evolution of the transport discipline. Secondly, based on the CoDa analysis, we find that the

development of the YMPS has gone through four stages and the YMPS is characterized by a tendency towards a "multi-core development" and faces a differentiated pattern of "peripheral port challenges". Thirdly, we take economic and policy factors into account to explains the underlying reason for the prosperity of the YMPS and provides a direction for its future development.

This paper is presented in the following structure: In Section 2, the YMPS is introduced. Section 3 describes the mathematical theory, including a concentration index (Normalized Herfindahl-Hirschman index (H^*)), the ternary diagram, and CoDa techniques. In Section 4, we apply CoDa exploratory tools (i.e., biplot and CoDa-dendrogram) to analyze the development pattern of the YMPS, looking for port associations and similarities from its temporal evolution. In Section 5, we discuss the results and provide some development experiences. Finally, Section 6 outlines the findings and makes considerations for future studies.

2 Overview of the Yangtze River Delta multi-port system (YMPS)

2.1 The description of the YMPS

The Yangtze River Delta region is the earliest pilot containerization area in China, which plays an exemplary role in the development and China's opening-up(Yang, Xiu, and Chen 2019). With the deep promotion of container technology and the gradual improvement of the regional distribution system, the YMPS has always been a significant port clustering area (Yang et al. 2019). From 1992 to 2019, the total container throughput in the YMPS always accounted for more than 30% of China, with an average annual growth rate of 11.6%.

The Yangtze River Delta region is the most developed area in China, attracting global attention (Feng et al. 2020). In the development of the YMPS, SH gradually lost its monopoly on container transportation. The Shanghai Province, Zhejiang Province and Jiangsu Province in China show a tripartite confrontation since 2012 (Feng, Grifoll, and Zheng 2019). SH has the benefit of geographical convenience, while NB has the advantage of shipping costs (Wang and Yeo 2019). NB and SH as dual-hub ports have shown a competitive situation in the YMPS (Feng, Grifoll, and Zheng 2019). SH is currently a traditional port that primarily offers freight handling and logistics facilities,

but it is increasingly evolving into an extensive maritime cluster that combines logistics and maritime services (Shi et al. 2020). With the development of port integration, Shanghai International Shipping Centre is increasingly transforming into a multi-core transport center (Zhou, Chen, and Shao 2017). The greater degree of regional port integration can lower handling fees and increase container throughput (Dong, Zheng, and Lee 2018). The YMPS is going through a regionalization mainly related to SH (Veenstra and Notteboom 2011). The port integration between Ningbo port and Zhoushan port (NB-ZS) since 2006 has greatly enhanced the competitiveness of NB-ZS. The spatial and temporal evolution of NB-ZS experienced from a feeder port to a hub port and traffic predictions reveal that in 2026 the container throughput in NB-ZS would reach nearly 49 million TEU that is equal to SH (Feng, Grifoll, and Zheng 2019).

2.2 Container throughput evolution of the Yangtze River Delta multi-port system (YMPS)

The earliest containerization of the YMPS began in the 1980s. Considering the data availability and integrity, we select the container throughput data of the YMPS from 1992 to 2019 as the research objectives. The data mainly from the China Ports Yearbook (1999-2019) and the Ministry of Transport of the People's Republic of China (https://www.mot.gov.cn/shuju/).

The container throughput evolution of the YMPS is illustrated in Figure 2 and Figure 10. Overall, container throughput of the YMPS maintained an increase from 1992 to 2019. After 2000, this growth momentum was seemed to be more powerful. After the impact of the 2008 financial crisis, the container throughput of the YMPS experienced a slight decline in 2009. After 2009, container throughput continued to maintain intense growth momentum.

3 Methodology

This paper proposes a method that combines Hierarchical Clustering based on the *clr*-transformation with CoDa exploratory tools (i.e., biplot and dendrogram) to investigate the temporal and spatial evolution of the YMPS from 1992 to 2019. Firstly, to identify the temporal characteristics of the YMPS from 1992 to 2019, we use Hierarchical Clustering to category the development of the YMPS into four stages. Then we add these four stages as temporal factors to CoDa exploratory tools so that

the points clustering in *clr*-biplot or boxplot can display four different colors. We can see that the four colored points clustering is highly consistent with Hierarchical Clustering. Finally, we can easily indicate a differentiated pattern other methods cannot distinguish from CoDa exploratory tools. At the same time, the results obtained from CoDa techniques can also be identified by H*, the ternary diagram and Aitchison distance. In this method, we can find the temporal and spatial characteristics and differentiated pattern simultaneously that other methods cannot meet. In this section, CoDa techniques, H* and the ternary diagram are introduced briefly.

3.1Compositional Data (CoDa) methodology

Compositional data is a part of the whole (i.e., frequency and percentage), conveying relative information in the ratios between its components (Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2015). The D-parts simplex is a group of positive vectors closed to constant k and denoted by:

$$S^D = \left\{ x = [x_1, x_2, \dots, x_D] : x_1 > 0, x_2 > 0, \dots, x_n > 0; \sum_{i=1}^{D} x_i = k \right\}. \quad \text{(Eq. 1)}$$

When all of its components are purely strictly positive numbers and only carry relative information the row vector $x = [x_1, x_2, \dots, x_D]$ is a D-parts compositional data (Aitchison 1982; Aitchison 1986; Pawlowsky-Glahn et al. 2015). The constant k is any purely positive number, is called the closure, usually 1 or 100. When $k = 1$, CoDa is proportional data, when $k = 100$, the CoDa is percentage data.

In this case, x_1 means the container traffic share of port one (a composition) of the YMPS in 1992, x_2 is the container traffic share of port two (a composition) in the YMPS in 1993 and so on…, k=1, D is equal to 28, represents 28-parts.

The Aitchison distance is the Simplex characterization like the distance in Euclidean geometry. The Aitchison distance and Norm are given by Eq. 2 and Eq.3 (Aitchison 1986).

Aitchison Distance between composition x and composition $y \in S^D$,

$$\mathrm{d}_a(x,y) = \sqrt{\frac{1}{2D}\sum_{i=1}^{D}\sum_{j=1}^{D}(\ln(\frac{x_i}{x_j}) - \ln(\frac{y_i}{y_j}))^2}. \quad \text{(Eq. 2)}$$

Norm of composition $x \in S^D$,

$$\|x\|_a = \sqrt{\frac{1}{2D}\sum_{i=1}^{D}\sum_{j=1}^{D}\left(\ln\frac{x_i}{x_j}\right)^2}. \quad \text{(Eq. 3)}$$

The standard statistical measures (e.g., Pearson correlation) is based on real space, when it used to CoDa, it can lead to spurious correlations and erroneous conclusions. Based on the Aitchison geometry, a new set of descriptive measures has been defined. The central tendency measurement of a compositional data set $\boldsymbol{X}$ is called the center:

$$\mathrm{cen}\,(\boldsymbol{X}) = \hat{g}_j = \boldsymbol{C}[\hat{g}_1, \hat{g}_2, \dots, \hat{g}_D], \quad \text{(Eq. 4)}$$

$$\hat{g}_j = \left(\prod_{i=1}^{n} x_{ij}\right)^{\frac{1}{n}}, i = 1,2,\dots,n.\ j = 1,2,\dots,D. \quad \text{(Eq. 5)}$$

Where $\boldsymbol{C}$ is the closure operator to constant k, $\hat{g}_j$ is the geometric mean of composition x, i is the parts order in the data set, j is the order of the composition. Dispersion in CoDa can be described by the variation matrix, initially defined by Aitchison as

$$\mathrm{T} = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1D} \\ t_{21} & t_{22} & \dots & t_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ t_{D1} & t_{D2} & t_{D3} & t_{DD} \end{pmatrix}, \quad \text{(Eq. 6)}$$

where $t_{ij} = \mathrm{var}(\ln\frac{x_i}{x_j})$ is the variance of the log-ratio of parts i and j.

The total variance is a measure of a compositional sample's global dispersion, given by

$$\text{totvar}[X] = \frac{1}{2D}\sum_{i,j=1}^{D}\text{var}(\ln\frac{x_i}{x_j}) = \frac{1}{2D}\sum_{i,j=1}^{D} t_{ij}\,, \qquad \text{(Eq. 7)}$$

There are generally two approaches to research CoDa. One is to work directly on the Simplex. Another is to formulate the compositions as log-ratio coordinates and then apply standard statistical methods to the log-ratios (in real space). Some transformations based on the log-ratio approach have been developed gradually, such as the additive log-ratio (*alr*) transformation, *clr*-transformation (Aitchison 1982) and the isometric log-ratio (*ilr*) transformation (Egozcue et al. 2003). The *clr*-transformation of a composition $x = [x_1, x_2, \ldots x_n]$ is

$$clr(x) = \left[\ln\frac{x_1}{g_{m(x)}}, \ln\frac{x_2}{g_{m(x)}}, \ldots, \ln\frac{x_D}{g_{m(x)}}\right], \qquad \text{(Eq. 8)}$$

where $g_i(x) = (\prod_{i=1}^{D} x_i)^{1/D}$ is the geometric mean of the parts.

The c*lr*-transformation is an operation for compositions in the Simplex are translated compositions into the real vector as an isometry. The *clr*-transformation of the *D*-part composition is a vector of coordinates of dimension $D-1$ through the *ilr*-transformation: $S^D \to R^{D-1}$. The composition is expressed as the coordinate of orthogonal basis, and the transformation is also isometry. The main criterion of choosing an orthogonal basis is that it should enhance the interpretability of coordinate representation. The particular cases that deserve our attention are related to the sequential binary partition (SBP) of the constituent vectors of the basis (Egozcue and Pawlowsky-Glahn 2005). The primary goal of the bases obtained from an SBP is to make it simple to interpret the composition according to the clustered sections selected at each level of the partition.

The balance is the normalized log-ratios of the geometric mean of the group of parts defined by the sign matrix at each step. In each balance, all parts are divided into two groups. This procedure is repeated until each group has just one part. For the $i\ th$ order partition, the definition of balance is described as follows: if the r parts $(i_1,\ i_2, \ldots, i_r)$ of the first subgroup are coded by +1 and the s parts

$(j_1, j_2, \ldots, j_s)$ of the second subgroup coded by -1 (see Table 2and Figure 7) (Egozcue and Pawlowsky-Glahn 2005). Balances are defined as follow:

$$b_k = \sqrt{\frac{rs}{r+s}} \ln \frac{(x_{i_1}, x_{i_2}, \ldots x_{i_r})^{\frac{1}{r}}}{(x_{j_1}, x_{j_2}, \ldots x_{j_r})^{\frac{1}{s}}} = \ln \frac{(x_{i_1}, x_{i_2}, \ldots x_{i_r})^{a_+}}{(x_{j_1}, x_{j_2}, \ldots x_{j_r})^{a_-}}, \quad \text{(Eq.9)}$$

where $a_+ = +\frac{1}{r}\sqrt{\frac{rs}{r+s}}$, $a_- = -\frac{1}{s}\sqrt{\frac{rs}{r+s}}$, and the values of r and s belong to the kth order partition, respectively.

3.2 Normalized Herfindahl-Hirschman index (H*)

Usually, H* is used to indicate market concentration (Elbayoumi and Dawood 2016), which is expressed by the formula:

$$H^* = \frac{\frac{\sum_{i=1}^{n} TEU_i^2}{\left(\sum_{i=1}^{n} TEU_i\right)^2} - \frac{1}{n}}{1 - \frac{1}{n}}, \quad \text{(Eq. 10)}$$

where i to n is the number of ports, TEU_i is the container throughput of port i. H* has a range of $[0, 1]$. When the market is a monopoly by one port, the H* is equal to 100%, if the H* is equal to 0, it means the market is divided equally by all the ports.

3.3 The ternary diagram

Three variables are depicted graphically as points in an equilateral triangle in the ternary diagram, it is also can describe the Aitchison distance of two points x and y in S^D (see Eq.2 and Figure 4). Viviani's theorem is the theoretical foundation of the ternary diagram (Abboud et al. 2010). The ternary diagrams have a broad range of applications in chemistry (Radivojević et al. 2018), geology (Promentilla et al. 2016), energy analysis (Petrik et al. 2018) and other fields. We divide the big equilateral triangle into four smaller equilateral triangles at the midpoints of the axes to help explain the concentration in the ternary diagram (see Figure 4) (Feng et al. 2020). The three small equilateral triangles on the corners are labeled "A Dominating", "B Dominating" and "C Dominating", while the middle triangle is labeled "Effective Competition" (Shepherd 2003). For instance, Port A has a market share of more than 50% in the "A-dominated" region. As a result, Port A dominates the industry.

242 However, in the “Effectively Competition” region, each port has less than a 50% market share, and
243 no single port can dominate the competition.

244 # 4 Results

245 Figure 3 illustrates the evolution of the container traffic share from 1992 to 2019. The sum of the
246 container traffic share of Shanghai (SH) and Ningbo (NB) always accounted for more than 80% in
247 the YMPS from 1992 - 2019. In 1998, SH’s traffic share reached the highest level (SH, NB and
248 OTH’s traffic share was 74.63%, 5.41% and 19.96%, respectively). After 1998, the container traffic
249 share of SH has been declining slowly until 2013, when the container traffic share of SH remained at
250 about 47%, however, NB’s container traffic share gradually increased since 1998 until to today. From
251 1992 to 2012, SH’s container traffic share accounted for more than 50% and monopolized the most
252 container traffic share. By 2012, SH’s traffic share had dropped below 50% (49.34%) firstly. In 2005,
253 the sum of container traffic share of SH and NB achieved the highest level (i.e., traffic share of SH,
254 NB and OTH was 68.45%, 21.3% and 10.25%, respectively), and after 2005, the container traffic
255 share of the other thirteen ports (OTH) began to increase steadily. As a result of the 2008 financial
256 crisis, all ports in the YMPS grew less than 0 in 2009. Other than that, all ports maintained strong
257 growth and the YMPS is in an enormous development momentum.

258 For 2019, SH’s traffic share reached the lowest level and NB’s traffic share reached the highest level,
259 the traffic share of SH, NB and OTH accounted for 46.95%, 29.85% and 23.2%, respectively.
260 Consequently, the development of the YMPS from 1992 to 2019 has experienced different stages (see
261 Figure 5). These stages are indicated by Hierarchical Clustering. Hierarchical clustering uses ward
262 criterion based on *clr*-transformation, which allows the extraction of information on temporal
263 evolution by defining similarity over the years. As example of *clr*-transformation, Table 1 include the
264 values of the log-ratio and the geometric mean for the year 2019. Using Hierarchical Clustering, we
265 categorize the container throughput evolution of the YMPS into four stages: 1992-1995, 1996-2000,
266 2001-2013and 2014-2019. Then we add the temporal factors (four stages) to CoDa exploratory tools
267 (*clr*-biplot and CoDa-dendrogram) to explore the temporal and spatial characteristics. The four stages
268 label the raw dataset into four groups. Their role is to color the “points” in the biplot or the “boxplots”

in CoDa-dendrogram so that can show different colored “points” in the biplot and different colored “boxplots” in CoDa-dendrogram (see Figure 6 and Figure 7). It also reflects that the temporal characteristics is highly consistent with H* and Aitchison distance (see Figure 9).

Principal Component Analysis is a multivariate statistical method that can be applied to analyze CoDa through a *clr*-transformation. The *clr*-biplot display multidimensional data points cloud by projecting data points into two-dimensional or at most three-dimensional space (see Figure 6) (Buccianti and Grunsky 2014). In the *clr*-biplot, some characteristics must be introduced, for example, a *clr*-biplot is composed of points, rays and links, and also by links formed by rays. In this case, the length of rays is proportional to variability, orthogonal connection means that the two sub-compositions are not related. Another explanation is that if the three rays are aligned, the relationship between the three parts is linear (Daunis-I-Estadella, Thió-Henestrosa, and Mateu-Figueras 2011; Pawlowsky-Glahn et al., 2015).

Table 1 shows the normalized variation matrix of the container throughput of the YMPS. The pairs give the largest contributions: NT-JX (5.98), JY-HZ (5.21), NT-JY (5.03) and other contributions that used to involve JX, JY, HZ and TZJ. The *clr*-biplot also confirm these results. Figure 6 presents the *clr*-biplot of the container traffic share in the YMPS. More than 86% of the projection is built on the first two components. The largest contribution of variability is given by the rays of JY, HZ, TZJ and JX, as already indicated in Table 1. Due to its small container throughput with minimal fluctuations, their fluctuation will be more significant for the same increase in capacity. For example, the container traffic growth rate of JY in 1995, 2005, 2009, 2014 and 2019 was 0. 00%, -4.71%, 48.4%, -56.6% and -4.2% respectively. But the growth rate of SH is 27.4%, 24.2%, -10.7%, 4.5% and 3% in 1995, 2005, 2009, 2014 and 2019 respectively, it is extremely stable compared with JY fluctuates in 2008 financial crisis. The other two larger variability are HZ and JX. HZ’s container traffic growth rate in 2009 and 2019 was 100% and -3.5%, respectively and JX was 0% and 6.5%.

According to Eq. 9, the compositions are expressed as balance 1 to balance 14 (b1, b2, ..., b14). The sequential binary partition is used to achieve each balance (SBP) is outlined in Table 2. Once the balances are received, the CoDa-dendrogram is a valuable tool to describe these balances. The CoDa-dendrogram is represented by the dendrogram-type links between groups, the horizontal lines do not

297 contain any information other than connected the group of parts, the length of the vertical lines shown
298 in Figure 7 is equal to the variance of each balance, and they represent the decomposition of the total
299 variance, the intersections of the horizontal line and vertical line determined the mean of a balance.
300 (Thió-Henestrosa et al. 2008; Pawlowsky-Glahn et al. 2015). The CoDa-dendrogram also shows the
301 *ilr* dispersion and quartiles through the boxplots. The first balance corresponds to the longest vertical
302 line, JX, TZJ, JY and HZ are separated from the geometric mean (center) of the remaining ports. As
303 the temporal element is added to the CoDa-dendrogram, the variances of the balances change in
304 Figure 7. In the first balance, the balance center has an obvious change than the other balance, which
305 means JX, TZJ, JY and HZ have an increasing influence compared to the other ports. This is also
306 reflected in the aforementioned *clr*-biplot (Figure 6) and the normalized variation matrix (Table 1).

307 In Figure 6, we refer to the ray of NJ, TZZ, NT, ZJ, SH, SZ, WZ, LYG and NB as group 1. We add
308 the temporal factor (four stages) to the *clr*-biplot, the clustering of the four colored points in Figure 6
309 means the four stages consistent with H* and Aitchison distance (see Figure 9). The links between
310 any rays in group 1 with JX or TZJ and the links between JY and HZ are approximately orthogonal.
311 It is implying that the log-ratios of the respective ports indicate a low correlation (i.e., the correlation
312 between ln (JX/ any port in group 1) and ln (JY/HZ) is -0.085). From a temporal perspective, JY has
313 an excellent development after 2000, but it gradually lost this advantage after 2010 (see Figure 6 and
314 Figure 7). This tendency can also be observed in Figure 2. From 2014 to 2019, the advantage of JX,
315 JY and HZ is gradually shifting to HZ.

316 Aitchison distance can be calculated by Eq.2. Figure 9 displays the evolution of H* and Aitchison
317 distance of the YMPS from 1992 to 2019, where we can see that Aitchison distance is strongly
318 consistent with H*. Simultaneously, the “turning points” in H* and Aitchison distance coincide with
319 the four stages of Hierarchical Clustering (i.e., 1995 and 2013). The peak of H* was in 1998 at 0.5
320 and the highest of Aitchison distance is 1.855 in 1992 respectively, then H* and Aitchison distance
321 all decreased to the lowest point in 2018 at 0.23 and 0.47 respectively. Since 2013, the points in the
322 ternary diagram have steadily shifted from SH to “Effective Competition” implying that SH has
323 gradually lost its monopoly status. This tendency is also evidenced in Figure 6 and Figure 7, for
324 example, from 2000 to 2010, points gradually shift from SH to JY, TZJ and JX. After 2014 this

tendency shifts to HZ. In Figure 7, in the first balance (b1), the third and fourth stages (2001-2013, 2014-2019) are closed to JX, JY, TZJ and HZ. In b12 balance, the development is focused on JY, TZJ and JX. The results are consistent with biplot in Figure 6. Meanwhile, H* and Aitchison distance remained at 0.24 and 0.5 with slight fluctuations after 2013.

5 Discussion

The benefits of CoDa techniques in transport disciplines are identified from the YMPS application results. In this way, we can simultaneously find temporal and spatial characteristics and distinguish a differentiated pattern (peripheral port challenges) that other methods cannot meet. The spatial evolution of multi-port systems may include several stages in the function of the opportunity factors such as infrastructure development, global trade tendency, shipping atmosphere or administrative issues, and the SH's monopolistic status as China's mainland gateway was no longer valid (Feng, Grifoll, and Zheng 2019). Simultaneously, ports within the YMPS use their strengths to enhance port-port and port-hinterland cooperation and extend its economic radius, making the port development a vital force for urban and regional development (Sakalayen, Chen, and Cahoon 2017). The spatial evolution of the YMPS continuous expansion, which roughly presents a northwest-southeast distribution pattern in space (see Figure 1 and Figure 10). The four stages of the YED multi-port system are characterizing as follow:

i.) Original single-core: 1992-1995

This first stage interpretation is connected with the excellent geographical location and advanced economy. Since the reform and opening-up in 1978, Pudong in Shanghai was established as an international financial center and shipping center. It performed most of the ocean transportation of the multi-port system and undertook the transshipment of other traditional port cargo in the multi-port system (Wang and Ducruet 2012). During this period, the traffic share was mainly concentrated in SH due to the expansion of global supply chains in containerization (see Figure 3) (Guerrero and Rodrigue 2014). The other ports, except for SH, were isolated and less connected from each other. Simultaneously, SH in the YMPS was relatively single in terms of landward, and the seaward is just

starting to take off. Consequently, other ports in the YMPS began to implement containerized transportation gradually, and the connection among ports in the YMPS became closer.

ii.) Polarization single-core: 1996-2000

The second period according to 1996-2000 is a process of polarization (see Figure 10). Under the effect of initial advantages (i.e., the excellent geographical location and policy support) and scale economy mechanism, SH integrated the import and export transportation and transshipment, attracting more container resources, the H* and Aitchison distance continued to rise until 2000. Simultaneously, the connection between the better-developed container ports and the inland cities was progressively strengthening. A minority of ports became sub-centers within the multi-port system (see Figure 1 and Figure 10).

iii.) Dual-core development: 2001-2013

The deconcentration shown in the ternary diagram, H* and Aitchison distance suggests that the YMPS began to change from neighboring ports. After 2001, The development center shifted to JX, TZJ, HZ and JY (see Figure 6) and the "peripheral ports challenges" appeared (see Figure 8, Figure 9 and Figure 10). Since China's accession to the WTO and opening-up policy had extended from coastal areas to inland in 2001, the port's capacity exceeded 10 million TEU from 2001 to 2005, with a utilization rate of 161%. In 2005, China established its first bonded port zone at Yangshan in Shanghai, which attracted Hong Kong's traffic share due to the comparable tariff advantages of a bonded port and a free trade port (Yang, Xiu, and Chen 2019).

With its excellent geographical advantage and increased cargo flow, the sub-core port (NB) has been positively expanding and exploring new shipping routes, especially ocean shipping routes (see Figure1 and Figure 10). As a result, many shipping enterprises transferred by sub-core port directly, which greatly stimulated the sub-core port development (see Figure 10). The sub-core port has gradually developed into a port comparable in function and scale to the core-hub port and has become another core-hub port of the YMPS (Feng, Grifoll, and Zheng 2019). Therefore, SH and NB's expansion accelerates the development of neighboring ports, a few ports have evolved from small ports to local hubs (i.e., JX and HZ). Simultaneously, due to enhancing the inland and sea collection

and distribution network, a common economic hinterland has emerged between NB and SH formed a transportation corridor. Affected by the 2008 financial crisis, throughput in the YDR multi-port system fell by 6.45% but grew by 13.33% in 2010 compared to 2008. After the 2008 financial crisis, all ports in the YMPS experienced negative container traffic growth (see Figure 2). However, SH and NB were much better able to adapt to the 2008 financial crisis and environmental impacts, and after 2009, all ports were back to the state they were before the 2008 financial crisis (see Figure 3).

iv.) Multi-core development: 2014-2019

With further development of the YMPS, the core-hub ports were beginning to be affected by the "peripheral ports challenges" mechanism (see Figure 10) (Wang et al. 2012). After 2013, ports in the YMPS have entered a new situation. Overall, the multi-port system appeared an evident dispersion. For example, the ports along the Yangtze River (i.e., NJ, SZ and NT) have initially built regional shipping centers. LYG is the intersection of "the Belt and the Road" (The 21st Century Maritime Silk Road on Land and Sea) and is a gateway to the Silk Road Economic Belt. At the same time, the container traffic share of JX, HZ, JY and TZJ have changed significantly (see Figure 3), and the "challenge" of small ports has become evident. Ports in the YMPS trended towards a "multi-core development" (Zhao et al. 2020) (see Figure 10). In this sense, CoDa has provided a robust tool to identify patterns when the magnitudes of the big ports particularly dominated the temporal evolution, for example, in Figure 2 the container throughput in SH port was 85 times larger than for JX port, the four small ports were isolated from other ports in Figure 7. JX and HZ showed a differentiated pattern thanks to their geographical position near SH and NB. In the context of China's foreign-oriented economy, a large number of goods supplied in the midstream and upstream of the Yangtze River needed to be transshipped to SH and NB by JX and HZ (see Figure 10). Consequently, it accelerated the development of JX and HZ. JY and TZJ belong to Jiangsu Province. Jiangsu Province ranks 2nd in China and 1st in the Yangtze River Delta region in terms of GDP, which indicates the ports in Jiangsu Province have huge hinterland and a vast amount of cargo need to be transshipped by ports.

Following the above analysis, the spatial and temporal evolution of traffic share is characterized by the "Original single-core"-"Dual-core development"-"Multi-core development" (see Figure 10). At the end of the 20th century, China built numerous deep-water terminals in SH and NB. As a result,

406 the container traffic share of SH decreased from 75% to 47% from 1992 to 2019, NB' traffic share
407 increased from 5.4% to 29% and OTH raised from 19.6% to 24% respectively from 1992 to 2019,
408 which demonstrated the "multi-core development" in the YMPS is strengthened (see Figure 1 and
409 Figure 10). Thus, NB and OTH posed a challenge to SH. The "challenge" effect was minimal and
410 mainly manifested in the challenge of the second largest port, NB, to the first largest port, SH. The
411 conclusions drawn from the CoDa techniques agrees with competitiveness analysis of NH and SH
412 postulated by Feng et al. (2020) and Gao and Li (2019) where hinterland, natural endowment and
413 services were identified as potential competition variables. Also, these ports had fallen into an
414 efficient competition taken shape in a twin-hub port.

415 This paper proposes a method that combines Hierarchical Clustering based on the *clr*-transformation
416 with CoDa exploratory tools (i.e., biplot and dendrogram) to investigate the temporal and spatial
417 evolution of the YMPS from 1992 to 2019. Firstly, to identify the temporal characteristics of the
418 YMPS from 1992 to 2019, we use Hierarchical Clustering to category the development of the YMPS
419 into four stages. Then we add these four stages as temporal factors to CoDa exploratory tools so that
420 the points clustering in *clr*-biplot or boxplot can display four different colors. We can see that the four
421 colored points clustering is highly consistent with Hierarchical Clustering. Finally, we can easily
422 indicate a differentiated pattern other methods cannot distinguish from CoDa exploratory tools. At
423 the same time, the results obtained from CoDa techniques can also be identified by H*, the ternary
424 diagram and Aitchison distance. In this method, we can find the temporal and spatial characteristics
425 and differentiated pattern simultaneously that other methods cannot meet.

426 Through the abovementioned analysis, CoDa techniques are an excellent way to explore the temporal
427 evolution and spatial integration in market share (Grifoll et al. 2019). Some concentration indexes
428 like Herfindahl-Hirschman Index (*H**) or Gini coefficients can only describe concentration or
429 deconcentration. However, CoDa techniques (ie., *clr*-biplot and CoDa-dendrogram) can explain the
430 temporal evolution and study the spatial characteristics simultaneously. CoDa techniques can also
431 find the differentiated development pattern that other methods cannot meet. At the same time, the
432 identification of the peripheral ports is also a good demonstration of the benefits of CoDa techniques.

6 Conclusions

As the YMPS plays an increasingly important role in international trade and the container traffic market, the YMPS has received more and more attention from scholars. In this paper, a method is proposed to investigate the temporal and spatial evolution of the YMPS from 1992 to 2019. The method can explore the temporal and spatial evolution simultaneously. To the best of my knowledge, this is the first approach to analyze the temporal and spatial characteristics simultaneously in the YMPS. Through the verification of traditional methods, such as H*, Aitchison distance and the ternary diagram, we get the following conclusions.

Firstly, in discipline, we propose a method that combines the Hierarchical Clustering with compositional data (CoDa) exploratory tools to explore the temporal and spatial evolution of the YMPS from 1992 to 2019. This method can simultaneously identify the temporal and spatial characteristics and find the differentiated development pattern that other methods cannot meet. In this sense, this method contributes further to improve the port system's evolutionary model and provides a new perspective for the temporal and spatial evolution of the transport discipline.

Secondly, based on the CoDa analysis, we find that the development of the YMPS has gone through four stages and the YMPS is characterized by a tendency towards a "multi-core development" and faces a differentiated pattern of "peripheral port challenges". That means SH acts as the center of the YMPS and faces a challenge and the main challenge is from the second largest port NB. SH and NB's expansion accelerates the development of neighboring ports and emerging smaller ports.

Thirdly, we take economic and policy factors into account to explains the underlying reason for the prosperity of the YMPS and provides a direction for its future development. Such work also benefits policymakers and stakeholders to make better decisions involving infrastructure management, business decision and resource allocation.

In addition, in the experimental case of the YMPS, we believe that CoDa techniques can apply to other multi-port regions, such as the Pearl River Delta multi-port system and the Bohai Rim multi-port system. From a development perspective, we have focused on a few relatively large ports, and

these ports are facing the challenges of small and medium-sized ports, so in future work, the internal competition of the multi-port region is attractive work.

References

Abboud, E. 2010. "Viviani's Theorem and Its Extension." The College Mathematics Journal 41 (3): 203–11. doi:10.4169/074683410X488683.

Aitchison, J. 1982. "The Statistical Analysis of Compositional Data." Journal of the Royal Statistical Society: Series B (Methodological) 44 (2): 139–60. doi:10.1111/j.2517-6161. 1982.tb01195. x.

Jones, M., J. Aitchison. 1986. "The Statistical Analysis of Compositional Data." Journal of the Royal Statistical Society. Series A (General) (1987) 150(4) 396. doi: 10.2307/2982045.

Baxter, M. J., and I. C. Freestone. 2006. "Log-Ratio Compositional Data Analysis in Archaeometry." Archaeometry 48 (July): 511–31. doi:10.1111/j.1475-4754.2006.00270. x.

Buccianti, A., and E. Grunsky. 2014. "Compositional Data Analysis in Geochemistry: Are We Sure to See What Really Occurs during Natural Processes?" Journal of Geochemical Exploration 141: 1–5. doi: 10.1016/j.gexplo.2014.03.022.

Cao, Y. H., Z. R. Jiang, S. L. Ye, W. Wu, and S. B. Liang. 2019. "Spatial Pattern and Heterogeneity of Port & Shipping Service Enterprises in the Yangtze River Delta, 2002–2016." Chinese Geographical Science 29 (3): 474–87. doi:10.1007/s11769-019-1035-1.

Christofakis, M., A. Tassopoulos, and B. Moukas. 2013. "Port Activity Evolution: The Initial Impact of Economic Crisis on Major Greek Ports." European Transport Research Review 5 (4): 195–205. doi:10.1007/s12544-013-0100-6.

Daunis-I-Estadella, J., S. Thió-Henestrosa, and G. Mateu-Figueras. 2011. "Two More Things about Compositional Biplot: Quality of Projection and Inclusion of Supplementary Elements." Proceedings of the 4th International Workshop on Compositional Data Analysis, 1–14.

Dong, G., S. Y. Zheng, and P. T. W. Lee. 2018. "The Effects of Regional Port Integration: The Case of NB-ZS." Transportation Research Part E: Logistics and Transportation Review 120 (1): 1–15. doi: 10.1016/j.tre.2018.10.008.

Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal. 2003. "Isometric Logratio Transformations for Compositional Data Analysis." Mathematical Geology 35 (3): 279–300. doi:10.1023/A:1023818214614.

Egozcue, J., and V Pawlowsky-Glahn. 2005. "CoDa-Dendrogram: A New Exploratory Tool." Compositional Data Analysis Workshop-CoDaWork'05, Proceedings, 1–10.

Elbayoumi, O., and A. Dawood. 2016. "Analysis of the Competition of Ports in the Middle East Container Ports Using HHI." Journal of Shipping and Ocean Engineering 6: 339–47. doi:10.17265/2159-5879/2016.06.003.

Feng H. X., M. Grifoll, P. J. Zheng, A. Martin-Mallofre, F. Murphy and S. Li. 2017. "Evolution and Container Traffic Prediction of Yangtze River Delta Multi-Port System (2001 – 2017)" International Journal of Shipping and Transport Logistics (January 2021). doi: 10.1504/IJSTL.2020.10024084

Feng, H. X., M. Grifoll, Z. Z. Yang, P. J. Zheng, and A. Martin-Mallofre. 2020. "Visualization of Container Throughput Evolution of the Yangtze River Delta Multi-Port System: The Ternary Diagram Method." Transportation Research Part E: Logistics and Transportation Review 142 (August 2019): 102039. doi: 10.1016/j.tre.2020.102039.

Feng, H. X., M. Grifoll, and P. J. Zheng. 2019a. "From a Feeder Port to a Hub Port: The Evolution Pathways, Dynamics

498 and Perspectives of NB-ZS (China).” Transport Policy 76 (August 2018): 21–35. doi:
499 10.1016/j.tranpol.2019.01.013.
500 Ferrer, B., G. Coenders, and E. Martínez-Garcia. 2015. “Determinants in Tourist Expenditure Composition — The Role
501 of Airline Types.” Tourism Economics 21 (1): 9–32. doi:10.5367/te.2014.0434.
502 Grifoll, M., M. I. Ortego, and J. J. Egozcue. 2019. “Compositional Data Techniques for the Analysis of the Container
503 Traffic Share in a Multi-Port Region.” European Transport Research Review 11 (1). doi:10.1186/s12544-019-0350-
504 z.
505 Grifoll, M., T. Karlis, and M. I. Ortego. 2018. “Characterizing the Evolution of the Container Traffic Share in the
506 Mediterranean Sea Using Hierarchical Clustering.” Journal of Marine Science and Engineering 6 (4).
507 doi:10.3390/jmse6040121.
508 Guerrero, D., and J. P. Rodrigue. 2014. “The Waves of Containerization: Shifts in Global Maritime Transportation.”
509 Journal of Transport Geography 34: 151–64. doi: 10.1016/j.jtrangeo.2013.12.003.
510 Notteboom, T. E. 1997. “Concentration and Load Centre Development in the European Container Port System.” Journal
511 of Transport Geography 5 (2): 99–115. doi:10.1016/s0966-6923(96)00072-5.
512 Pallis, A., and G. K. Vaggelas. 2017. “A Greek Prototype of Port Governance.” Research in Transportation Business and
513 Management 22: 49–57. doi: 10.1016/j.rtbm.2016.12.003.
514 Pawlowsky-Glahn, V., J. J. Egozcue, and R. Tolosana-Delgado. 2015. Modeling and Analysis of Compositional Data.
515 John Wiley & Sons, Ltd.: Chichester, UK, 2015; p. 253.
516 Petrik, A., S. Albanese, A. Lima, and B. D. Vivo. 2018. “The Spatial Pattern of Beryllium and Its Possible Origin Using
517 Compositional Data Analysis on a High-Density Topsoil Data Set from the Campania Region (Italy).” Applied
518 Geochemistry 91 (February): 162–73. doi: 10.1016/j.apgeochem.2018.02.008.
519 Promentilla, M. A. B, N. H. Thang, P. T. Kien, H. Hinode, F. T. Bacani, and S. M. Gallardo. 2016. “Optimizing Ternary-
520 Blended Geopolymers with Multi-Response Surface Analysis.” Waste and Biomass Valorization 7 (4): 929–39. doi:
521 10.1007/s12649-016-9490-8.
522 Radivojević, M., J. Pendić, A. Srejić, M. Korać, C. Davey, A. Benzonelli, M. Martinon-Torres, N. Jovanović, and Ž.
523 Kamberović. 2018. “Experimental Design of the Cu-As-Sn Ternary Colour Diagram.” Journal of Archaeological
524 Science 90: 106–19. doi: 10.1016/j.jas.2017.12.001.
525 Reimann, C., P. Filzmoser, K. Fabian, K. Hron, M. Birke, A. Demetriades, E. Dinelli, et al. 2012. “The Concept of
526 Compositional Data Analysis in Practice - Total Major Element Concentrations in Agricultural and Grazing Land
527 Soils of Europe.” Science of the Total Environment 426: 196–210. doi: 10.1016/j.scitotenv.2012.02.032.
528 Sakalayen, Q., P. S.-L. Chen, and S. Cahoon. 2017. “The Strategic Role of Ports in Regional Development:
529 Conceptualising the Experience from Australia.” Maritime Policy and Management 44 (8): 933–55.
530 doi:10.1080/03088839.2017.1367969.
531 Shepherd, W. G., and J. M. Shepherd. 2003. The Economics of Industrial Organization.
532 Shi, X., H. Z. Jiang, H. Li, and Y. Wang. 2020. “Upgrading Port-Originated Maritime Clusters: Insights from Shanghai’s
533 Experience.” Transport Policy 87 (December 2018): 19–32. doi: 10.1016/j.tranpol.2019.11.002.
534 Svindland, M., J. Monios, and H. M. Hjelle. 2019. “Port Rationalization and the Evolution of Regional Port Systems: The
535 Case of Norway.” Maritime Policy and Management 46 (5): 613–29. doi:10.1080/03088839.2019.1574988.
536 Thió-Henestrosa, S., J. J. Egozcue, V. Pawlowsky-Glahn, L. Ó Kovács, and G. P. Kovács. 2008. “Balance-Dendrogram.
537 A New Routine of CoDaPack.” Computers and Geosciences 34 (12): 1682–96.
538 https://doi.org/10.1016/j.cageo.2007.06.011.
539 Veenstra, A., and T. Notteboom. 2011. “The Development of the Yangtze River Container Port System.” Journal of
540 Transport Geography 19 (4): 772–81. doi: 10.1016/j.jtrangeo.2010.09.006.
541 Villa, J. C. 2017. “Port Reform in Mexico: 1993–2015.” Research in Transportation Business and Management 22: 232–

38. doi: 10.1016/j.rtbm.2016.11.003.
Wang, C. J., and C. Ducruet. 2012. “New Port Development and Global City Making: Emergence of the Shanghai-Yangshan Multilayered Gateway Hub.” Journal of Transport Geography 25 (November 2012): 58–69. doi: 10.1016/j.jtrangeo.2012.07.008.
Wang, C. J., J. E. Wang, and César Ducruet. 2012. “Peripheral Challenge in Container Port System: A Case Study of Pearl River Delta.” Chinese Geographical Science 22 (1): 97–108. doi:10.1007/s11769-012-0517-1.
Wang, Y., and G. T. Yeo. 2019. “Transshipment Hub Port Selection for Shipping Carriers in a Dual Hub-Port System.” Maritime Policy and Management 46 (6): 701–14. doi:10.1080/03088839.2019.1627012.
Yang, Z. Z., Q. H. Xiu, and D. X. Chen. 2019. “Historical Changes in the Port and Shipping Industry in Hong Kong and the Underlying Policies.” Transport Policy 82 (September 2017): 138–47. doi: 10.1016/j.tranpol.2018.03.007.
Zhao, D., Z. F. Li, Y. T. Zhou, X. Chen, and S. S. Liang. 2020. “Measurement and Spatial Spillover Effects of Port Comprehensive Strength: Empirical Evidence from China.” Transport Policy 99 (December 2019): 288–98. doi: 10.1016/j.tranpol.2020.09.006.
Zhoua, X., X. H. Chen, and L. Shao. 2017. “Study on the Optimization of Collection and Distribution System of Freight Hub Ports: Illustrated by the Case of Shanghai International Shipping Center, China.” Transportation Research Procedia 25: 1126–36. doi: 10.1016/j.trpro.2017.05.125.

571

572

573 **Table. 1** Normalized variation matrix of the traffic throughput yearly compositions and *clr*-transformation for the year
574 2019. The geometric mean for this year is equal to x.xx.

Port	SH	NB	SZ	NT	NJ	LYG	JY	TZJ	YZ	ZJ	TZZ	WZ	JX	HZ	HZH	*clr*
SH	0	0.35	0.32	0.17	0.15	0.45	3.56	1.98	0.11	0.14	0.17	0.07	4.83	4.54	1.22	
NB	0.35	0	0.23	0.84	0.55	0.16	1.84	0.70	0.18	0.44	0.59	0.24	2.71	3.37	1.06	
SZ	0.32	0.23	0	0.59	0.23	0.09	2.62	1.41	0.22	0.13	0.37	0.15	3.13	3.26	1.32	
NT	0.17	0.84	0.59	0	0.12	0.88	5.03	3.00	0.37	0.21	0.10	0.25	5.98	4.58	1.20	
NJ	0.15	0.55	0.23	0.12	0	0.45	4.14	2.39	0.22	0.03	0.09	0.09	4.73	3.81	1.24	
LYG	0.45	0.16	0.09	0.88	0.45	0	2.05	1.09	0.24	0.30	0.59	0.22	2.64	3.34	1.51	
JY	3.56	1.84	2.62	5.03	4.14	2.05	0	0.59	2.83	3.67	4.23	3.18	1.24	5.21	4.24	
TZJ	1.98	0.70	1.41	3.00	2.39	1.09	0.59	0	1.47	2.15	2.42	1.73	1.19	3.19	1.92	
YZ	0.11	0.18	0.22	0.37	0.22	0.24	2.83	1.47	0	0.17	0.25	0.07	3.87	3.78	1.09	
ZJ	0.14	0.44	0.13	0.21	0.03	0.30	3.67	2.15	0.17	0	0.12	0.06	4.33	3.84	1.33	
TZZ	0.17	0.59	0.37	0.10	0.09	0.59	4.23	2.42	0.25	0.12	0	0.15	4.92	3.78	0.97	
WZ	0.07	0.24	0.15	0.25	0.09	0.22	3.18	1.73	0.07	0.06	0.15	0	4.04	3.89	1.18	
JX	4.83	2.71	3.13	5.98	4.73	2.64	1.24	1.19	3.87	4.33	4.92	4.04	0	2.57	4.13	
HZ	4.54	3.37	3.26	4.58	3.81	3.34	5.21	3.19	3.78	3.84	3.78	3.89	2.57	0	2.36	
HZH	1.22	1.06	1.32	1.20	1.24	1.51	4.24	1.92	1.09	1.33	0.97	1.18	4.13	2.36	0	

575 **Table. 2** Sequential Binary Partition for the YMPS.

Balance	SH	NB	SZ	NT	NJ	LYG	JY	TZJ	YZ	ZJ	TZZ	WZ	JX	HZ	HZH
b1	1	1	1	1	1	1	-1	-1	1	1	1	1	-1	-1	1
b2	0	0	0	0	0	0	-1	-1	0	0	0	0	-1	1	0
b3	0	0	0	0	0	0	1	-1	0	0	0	0	-1	0	0
b4	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	0	0	1
b5	0	0	0	0	0	0	0	1	0	0	0	0	-1	0	0
b6	-1	1	1	-1	-1	1	0	0	-1	-1	-1	-1	0	0	0
b7	-1	0	0	-1	-1	0	0	0	1	-1	-1	-1	0	0	0
b8	-1	0	0	1	-1	0	0	0	0	-1	-1	-1	0	0	0
b9	0	1	-1	0	0	-1	0	0	0	0	0	0	0	0	0
b10	-1	0	0	0	1	0	0	0	0	-1	1	-1	0	0	0
b11	0	0	1	0	0	-1	0	0	0	0	0	0	0	0	0
b12	1	0	0	0	0	0	0	0	0	-1	0	-1	0	0	0
b13	0	0	0	0	1	0	0	0	0	0	-1	0	0	0	0
b14	0	0	0	0	0	0	0	0	0	1	0	-1	0	0	0

576

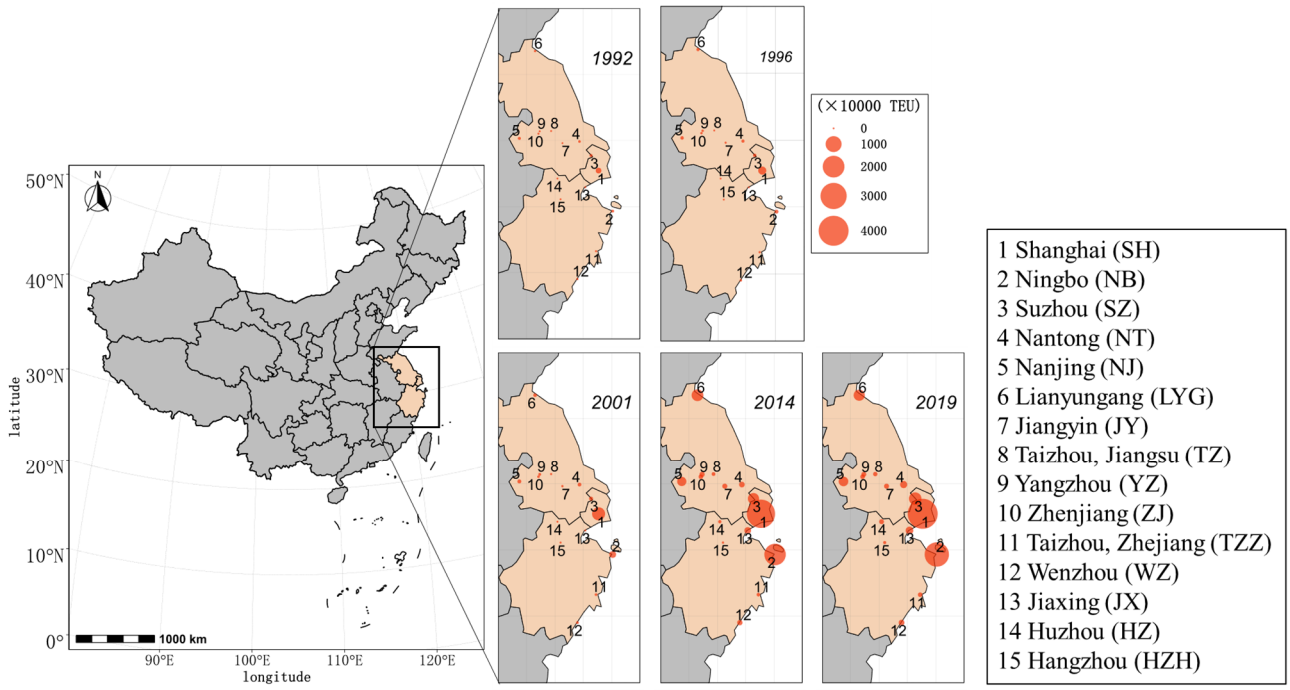
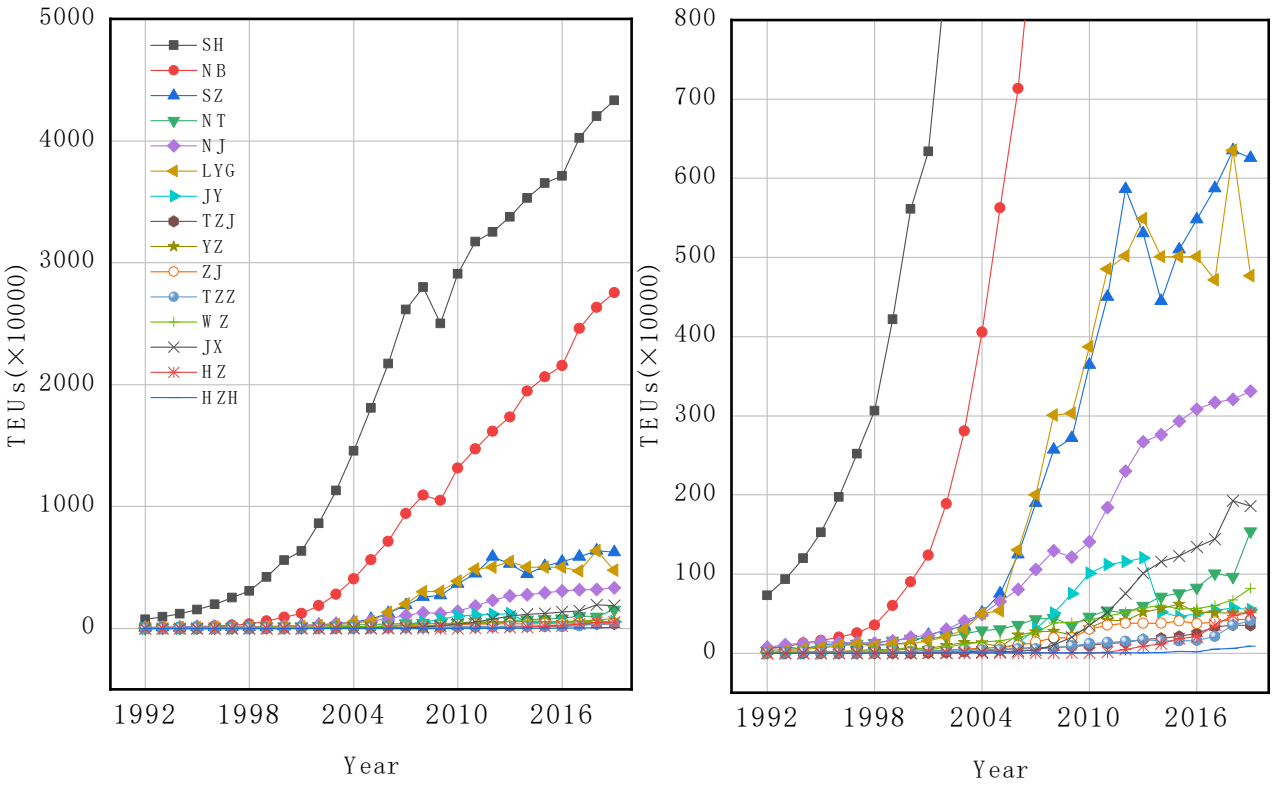


Figure.1 The location of the ports in the YMPS and the evolution of container throughput (in 10000 TEU) of the YMPS in 1992, 1996, 2001, 2014 and 2019. Ports are identified as follow: Shanghai port (SH), Suzhou port (SZ), Nantong port (NT), Nanjing port (NJ), Lianyungang port (LYG), Jiangyin port (JY), Taizhou port (Jiangsu Province, TZJ), Yangzhou port (YZ), Zhenjiang port (ZJ), Ningbo port (NB), Taizhou port (Zhejiang Province, TZZ), Wenzhou port (WZ), Jiaxing port (JX), Huzhou port (HZ) and Hangzhou port (HZH).



583

584 **Figure. 2** (left) Evolution of the container throughput in the YMPS. (right) Zoom of the lower range of the container
585 throughput.

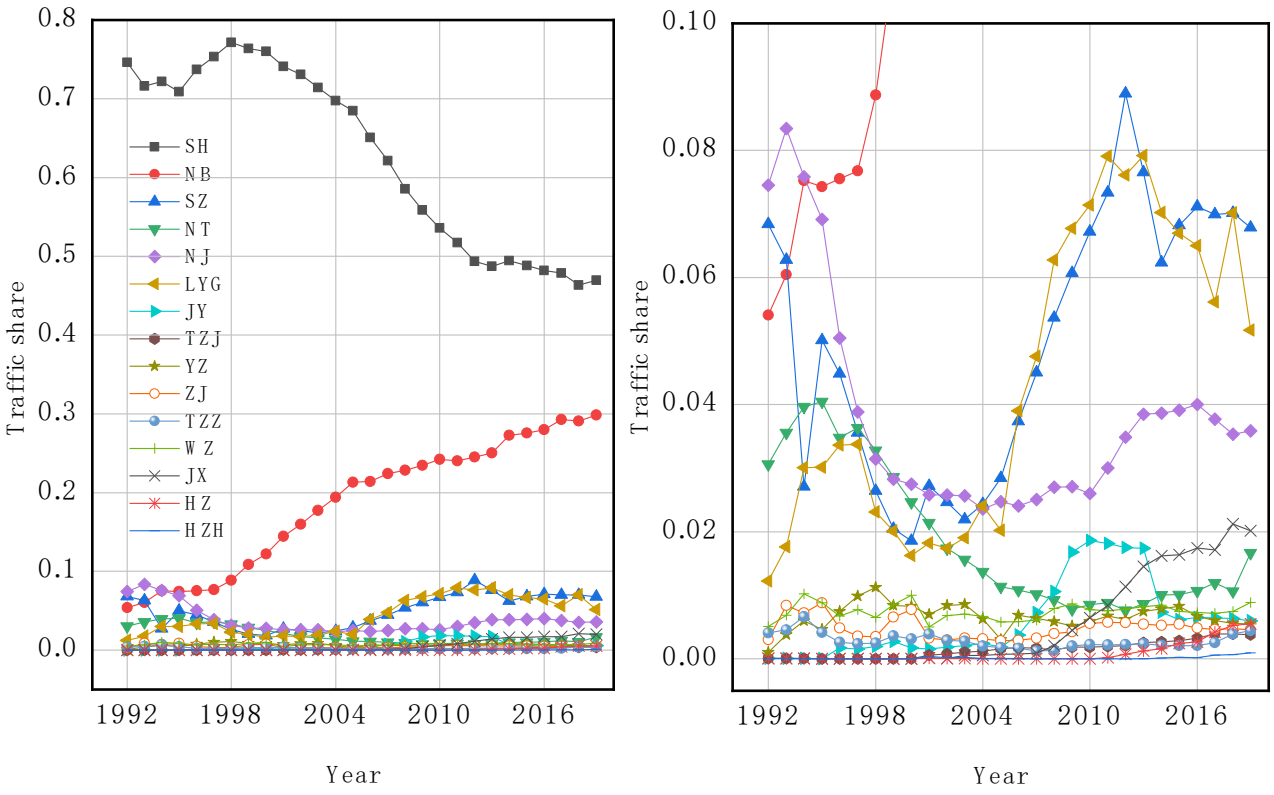


Figure. 3 (Left) Evolution of the container traffic share in the YMPS. (right) Zoom of the lower range of the container traffic share. Different colors represent the traffic share of different ports.

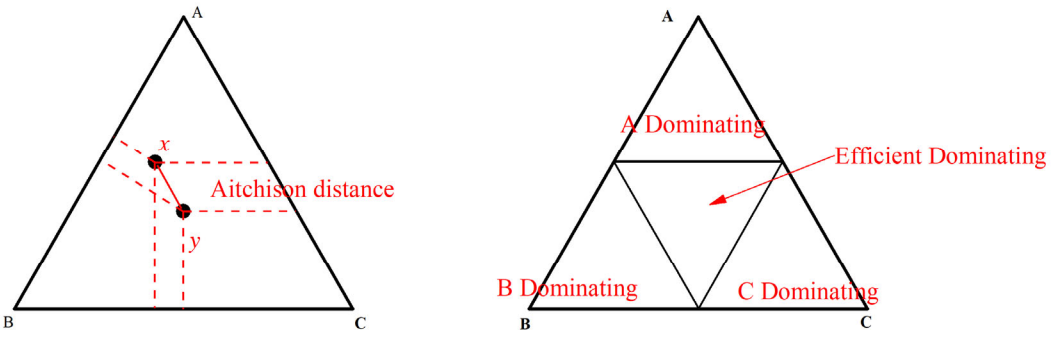


Figure. 4 (left) The Aitchison distance describe in the ternary. (right) Three parts composition data (A, B and C) are represented in the ternary diagram. Domain division of the ternary diagram and the container traffic share of SH, OTH and NB express in ternary diagram.

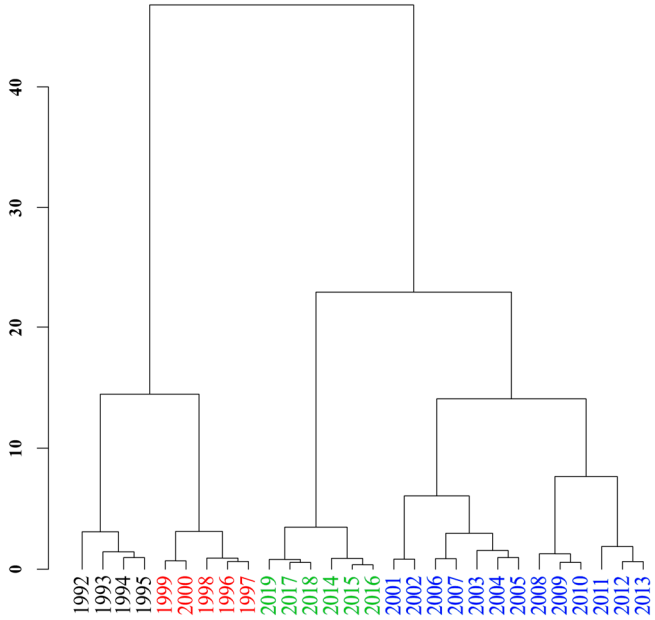


Figure. 5 Dendrogram (Hierarchical Clustering) for *clr*-transformed container traffic share of the YMPS from 1992 to 2019. Different colors represent different stages, the first period according to 1992-1995 (black), the second period is 1996-2000 (red), the third is 2001-2013 (blue) and the last is 2014-2019 (green).

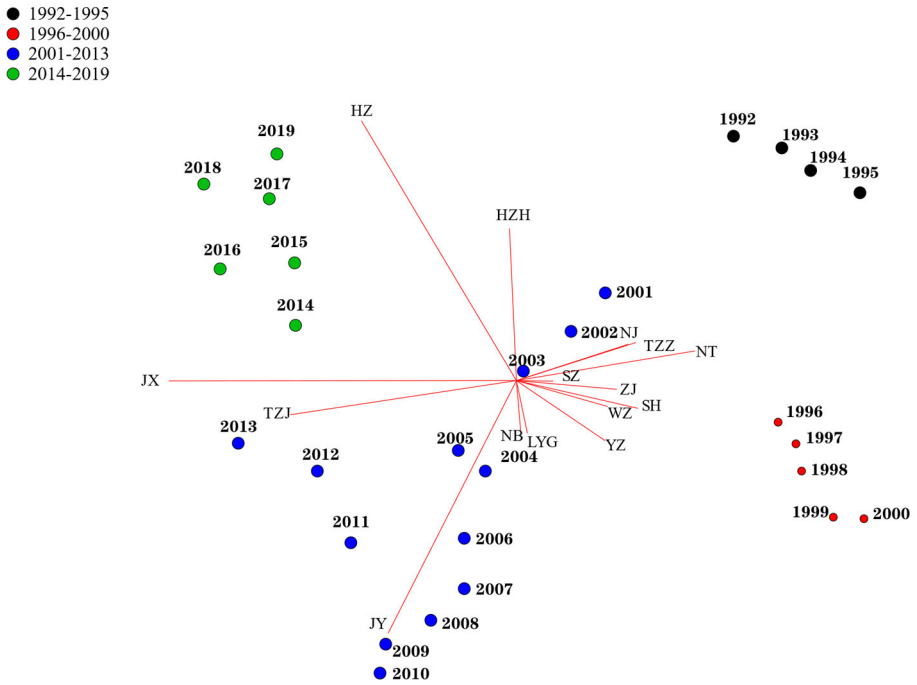
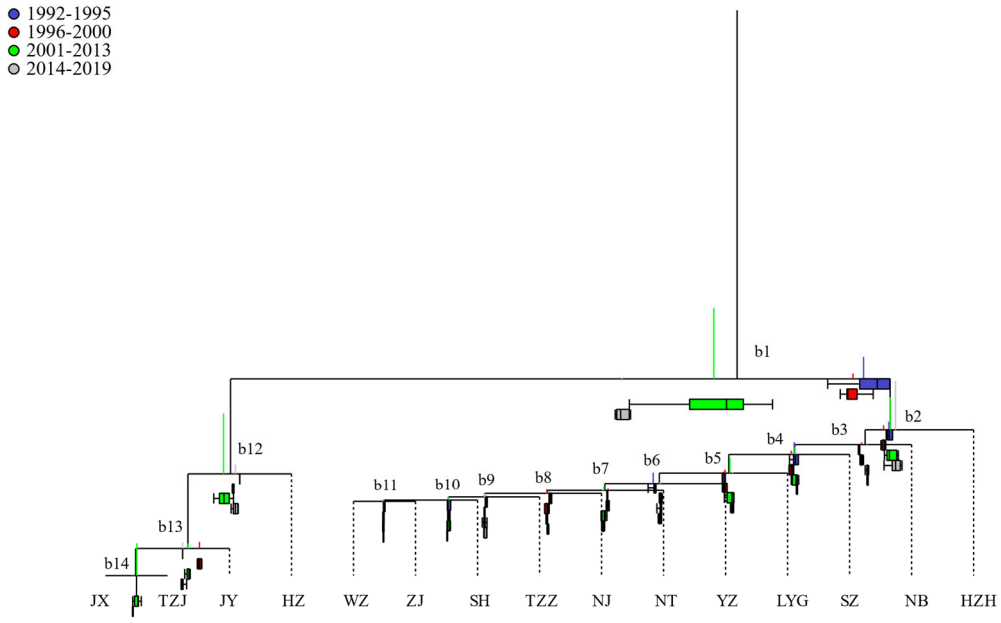
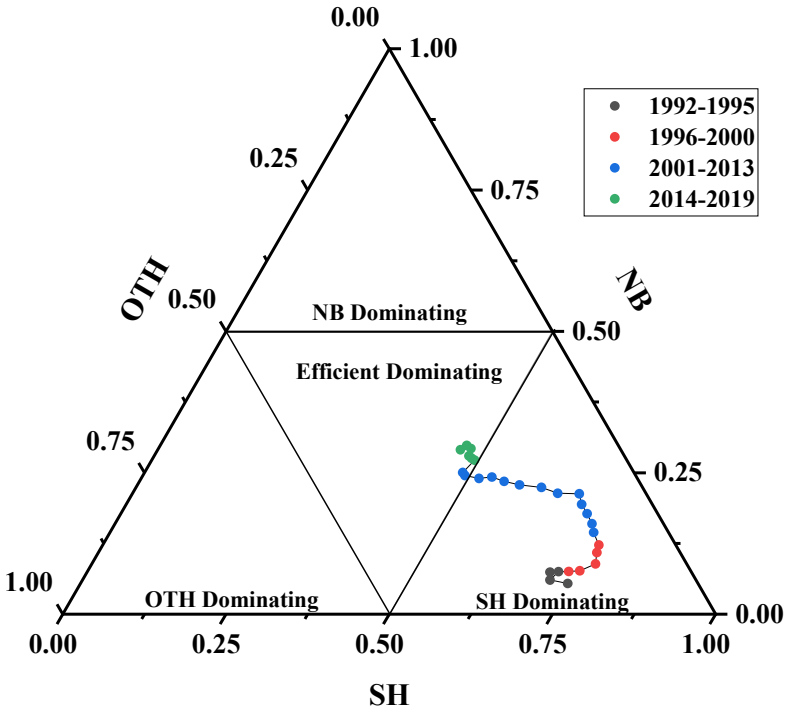


Figure. 6 The c*lr*-biplot, different coloured points represent different stages.



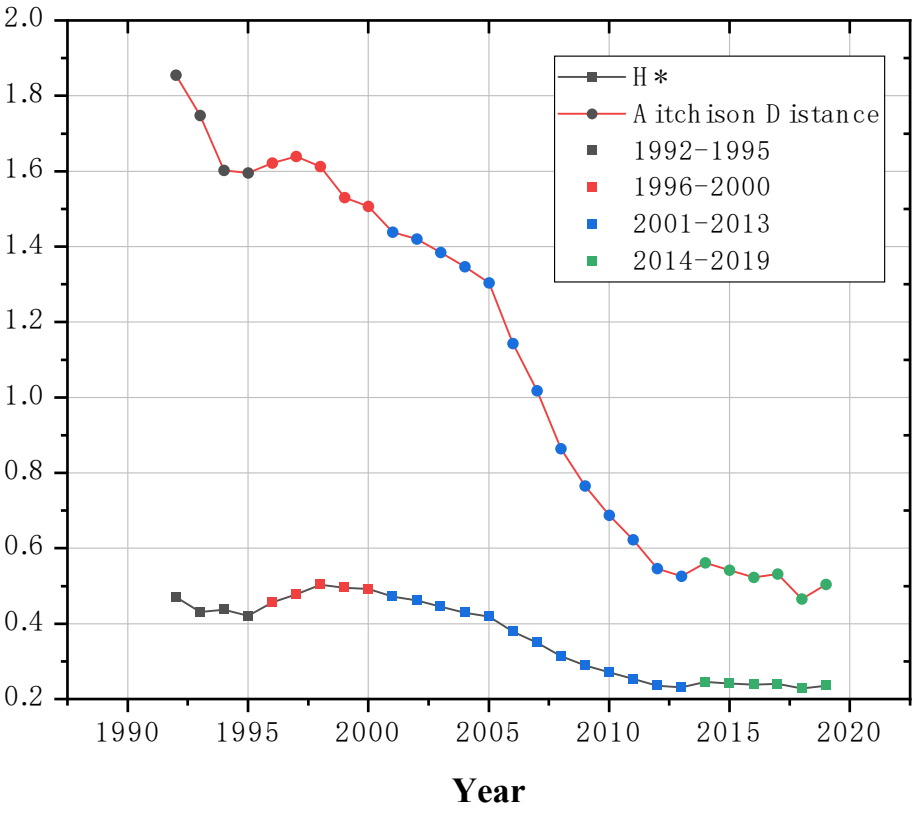
600

601 **Figure. 7** CoDa-dendrogram of the YMPS using the balances of Table 2.



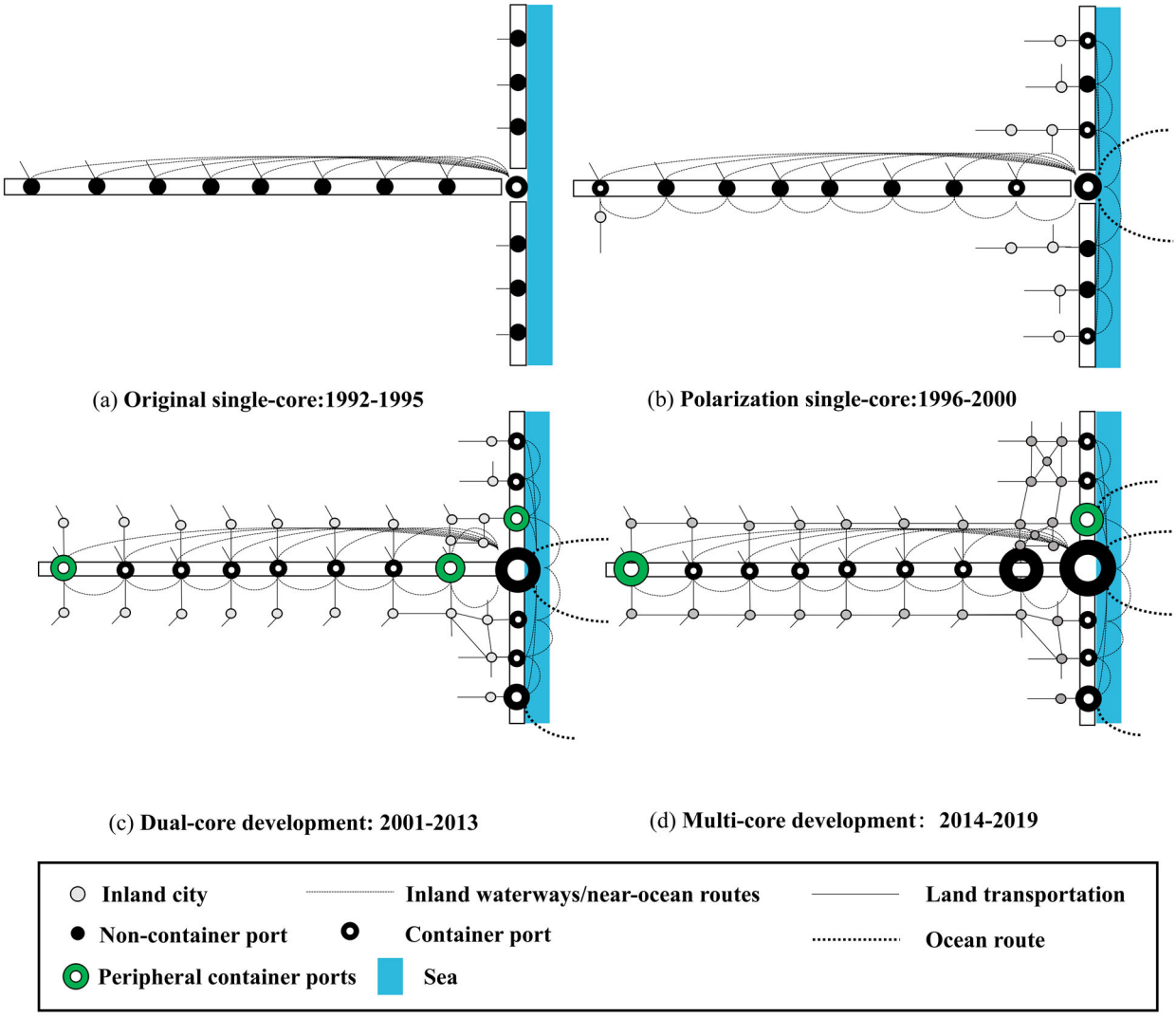
602

603 **Figure. 8** The YMPS container traffic share of SH, NB and OTH describe in the ternary diagram.



604

605 **Figure. 9** The evolution of H* and Aitchison distance for the YMPS from 1992 to 2019.



606

607 **Figure. 10** The temporal and spatial evolution model of the YMPS, the size of the circle represents the size of the port.

608