Gini Coefficient, Dissimilarity Index and Lorenz Curve for the Spanish Port System by type of goods

Nicoletta González-Cancelas

Civil Engineering. Transport
Universidad Politécnica de Madrid
Madrid, Spain
nicoleta.gcancelas@upm.es

Francisco Soler-Flores

Civil Engineering. Transport
Universidad Politécnica de Madrid
Madrid, Spain
fsoler@upm.esç

M. Carmen Palomino-Monzón

Departamento Ingenieria Civil: Ordenación Del Territorio,
Urbanismo y Medio Ambiente
Universidad Politécnica de Madrid
Madrid, Spain
mcpalomino@caminos.upm.es

J.Luis Almazán-Gárate

Departamento Ingenieria Civil: Ordenacion Del Territorio
Universidad Politécnica de Madrid
Madrid, Spain
joseluis.almazan@upm.es

Abstract— **This paper shows the Gini Coefficient, the dissimilarity Index and the Lorenz Curve for the Spanish Port System by type of goods from 1960 to the year 2010 for business units: Total traffic, Liquid bulk cargo, Solid bulk cargo, General Merchandise and Container (TEUs) with the aim of carcaterizar the Spanish port systems in these periods and propose future strategies.**

Key words-Spanish Port System Gini Coefficient, Dissimilarity Index Lorenz Curve

I.   INTRODUCTION

One of the main issues in port logistics or other related freight transport engineering fields is the general forecast of those parameters related to space, means, and resources requirements, as well as their optimisation. Physical and equipment parameters related to a container terminal (i.e. stocking surface, necessary berthing length, dock cranes number,...) represent a high investment and are characterised by important social, economical or environmental impacts.

Therefore, a correct forecast of these parameters and of the actual surface requirements (least possible geographical impact, thus its least modification), leads the performed research to provide a highly useful tool to any planning agent, so it can anticipate and/or forecast its space and means needs, way before strategic, marketing or planning decision making.

Up to this date, port planning has been rather based on empirical, analytical or simulation models. Empirical methods are based on productivity average indicators issued by planning agents. These indicators set a relationship between the main activities of a subsystem and the total annual production. These methods are thus very useful when dealing with new terminals planning or master plans development. The reference indicators have been constantly studied and updated by different authors over the years [1], [2], [3], [4], [5], [7], [8]. Analytical methods use mathematical concepts and formulas, based on the queuing theory and requiring large databases. These methods have been studied by several authors ([1], [2], [9]). The study

of [10] emphasized it in his paper "Port and container terminals modelling". The paper mention several studies (i.e.[11], [12], [13]), based on different aspects of the berthing system planning, as the occupation ratio, port congestion percentage, minimum waiting time, total port system costs, optimal number of berthing points and dock cranes, the optimal ratios berthing points/terminal or dock cranes/berthing points, etc. As indicated by UNCTAD[1] [2], simulation techniques use models to represent complex processes, whose mathematical description is not performable due to random behaviour and non-linear characteristics of the process. A detailed description of the method and the results of its application to the Casablanca Port is included in a paper published by UNCTAD [14].

The USA, [15] published a paper that performs a revision on the literature related to the capacity factors, focused on port planning. Other paper has been issued in Singapore [16] dealing with strategic planning issues.

Spanish bibliographical references start back in 1977, [1], with a paper stating the basics of port planning. And [3] would publish later on a comparison between exploitation conditions in several Spanish ports, using empirical methods. More recently, paper [17] presents the parameters and processes to be considered in a container terminal planning. In 2007, in his PhD thesis [18], determinates the characteristic parameters and ratios of the port operation, obtaining their values for each container port terminal. Other papers on logistical planning could also be mentioned [1].

The Artificial Intelligence, concretely the neural networks, is meant to significantly improve the ports' planning. There are practically no papers on neural networks' application in transports' planning, mainly because of the Artificial Intelligence recent release. The origins are set back in 1943 [19], facing a rather difficult start and lack of interest among

---

[1]   UNCTAD: United Nations Conference on Trade and Development (www.unctad.org)

the researchers. Higher interest for the artificial intelligence has been shown since 1982, when John Hopfield [20] has stated the Backpropagation algorithm.

Nowadays, well known universities (Boston, Helsinki, Stanford, Carnegie-Mellon, California, and Massachusetts) are developing research programs on neural networks, as well as some private societies in Japan, USA or Europe.

Neural networks' application in transports planning is illustrated in a paper published by Cadiz University, dealing with forecast techniques in road traffic [21].

More recent studies (2010) perform a container traffic forecast in Bangkok Port, using neural networks to explore their applicability to predict future container traffic needs and – through this – to estimate future investments in port extensions [22]. Other studies perform a comparison between traditional and neural networks based forecast techniques used to predict container traffic in the same port [23].

Other studies are issued in 2011 [24], analysing the advantages and/or differences between purely statistical methods and neural networks, in terms of transport research. Cited study deals with the particular suitability of the neural networks to represent non-linear phenomena and with their learning capacity.

There are also papers dealing with neural networks application in short term planning processes; these techniques have been applied to traffic parameters (i.e. flux or occupation) prediction [25], to traffic flux, speed and occupation [26], transportation general problems [27], [28] or to short term train passengers demand [29]. All of them have produced reliable results and promising feedbacks for the future use of the neural networks.

II. GINI COEFFICIENT, DISSIMILARITY INDEX AND LORENZ CURVE

The Gini coefficient was developed to measure the degree of concentration (inequality) of a variable in a distribution of its elements. It compares the Lorenz curve (figure 1) of a ranked empirical distribution with the line of perfect equality. This line assumes that each element has the same contribution to the total summation of the values of a variable. The Gini coefficient ranges between 0, where there is no concentration (perfect equality), and 1 where there is total concentration (perfect inequality).

The Lorenz curve is a graphical representation of the proportionality of a distribution (the cumulative percentage of the values). To build the Lorenz curve, all the elements of a distribution must be ordered from the most important to the least important. Then, each element is plotted according to their cumulative percentage of X and Y, X being the cumulative percentage of elements and Y being their cumulative importance. For instance, out of a distribution of 10 elements (N), the first element would represent 10% of X and whatever percentage of Y it represents (this percentage must be the highest in the distribution). The second element would cumulatively represent 20% of X (its 10% plus the 10% of the first element) and its percentage of Y plus the percentage of Y of the first element.
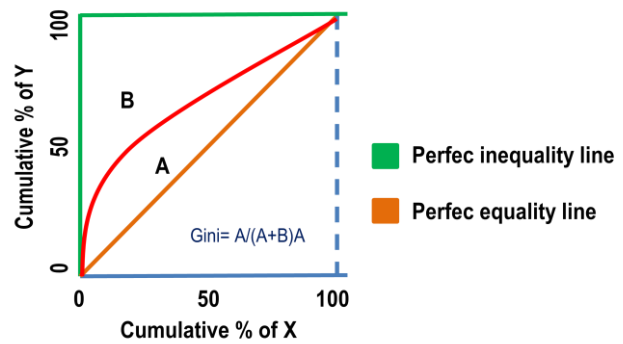

Figure 1.   The Lorenz Curve

The Lorenz curve is compared with the perfect equality line, which is a linear relationships that plots a distribution where each element has an equal value in its shares of X and Y. For instance, in a distribution of 10 elements, if there is perfect equality, the 5th element would have a cumulative percentage of 50% for X and Y. The perfect inequality line represents a distribution where one element has the total cumulative percentage of Y while the others have none.

The Gini coefficient is defined graphically as a ratio of two surfaces involving the summation of all vertical deviations between the Lorenz curve and the perfect equality line (A) divided by the difference between the perfect equality and perfect inequality lines (A+B).

Geographers and many others have used the Gini coefficient in numerous instances, such assessing income distribution among a set of contiguous regions (or countries) or to measure other spatial phenomena such industrial location. Its major purpose as a method in transport geography has been related to measuring the concentration of traffic (figure 1), mainly at terminals, such as assessing changes in port system concentration. Economies of scale in transportation can favor the concentration of traffic at transport terminals, while other considerations such as accessibility to regional markets can be perceived as a countervailing force to concentration. So, the temporal variations of the Gini coefficient reflect changes in the comparative advantages of a location within the transport system.
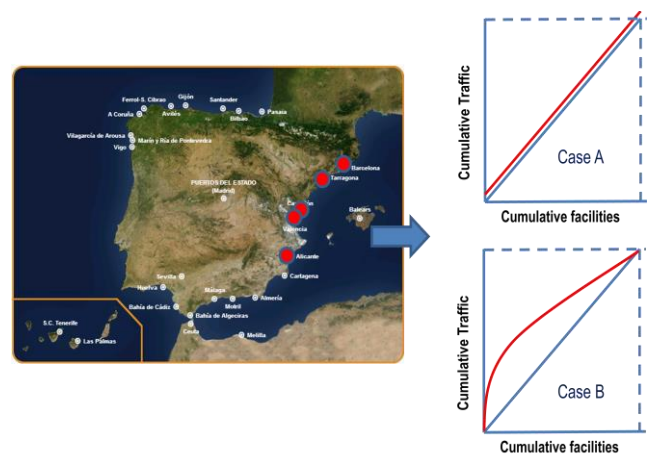

Figure 2.   Traffic Concentration and Lorenz Curves

The figure 2 represents a simple system of 5 ports along a coast. In case A, the traffic for each port is the same, so there is no concentration and thus no inequality. The Lorenz curve of this distribution is the same than the perfect equality line; they overlap. In case B, there is some concentration of the traffic in two ports and this concentration is reflected in the Lorenz curve as it is different from the perfect equality line. Case C represents a high level of concentration in *two ports* (for example Barcelona and Valencia) and the Lorenz curve is significantly different from the perfect equality line.

The dissimilarity index is the summation of vertical deviations between the Lorenz curve and the line of perfect equality, also known as the summation of Lorenz differences. The closer the ID is to 1 (or 100 if percentages are used instead of fractions), the more dissimilar the distribution is to the line of perfect equality.

$$ID = 0.5 \sum_{i=1}^{N} |X_i - Y_i| \tag{1}$$

Where X and Y are percentages (or fractions) of the total number of elements and their respective values (traffic being the most common). N is the number of elements (observations).

The Gini Coefficient represents the area of concentration between the Lorenz curve and the line of perfect equality as it expresses a proportion of the area enclosed by the triangle defined by the line of perfect equality and the line of perfect inequality. The closer the coefficient is to 1, the more unequal the distribution.

$$G = 1 - \sum_{i=0}^{N} \left(\sigma Y_{i-1} + \sigma Y_i \right) \left(\sigma X_{i-1} - \sigma X_i \right) \tag{2}$$

Where σX and σY are cumulative percentages of Xs and Ys (in fractions) and N is the number of elements (observations).

III. CALCULATION OF THE GINI COEFFICIENT, THE DISSIMILARITY INDEX AND THE LORENZ CURVE FOR THE SPANISH PORT SYSTEM BY TYPE OF GOODS

In order to characterize the Spanish Port System calculates the Gini coefficient, the dissimilarity index and Lorenz curve from the years 1960-2010 for the following business units (table I).

TABLE I.   BUSINESS UNITS

Business units
• Total traffic
• Liquid bulk cargo
• Solid bulk cargo
• General Merchandise
• Container (TEUs)
• Fresh fish
• Provisioning
• Interior traffic
• Ships

In this paper we will present only the results for: Total traffic, Liquid bulk cargo, Solid bulk cargo, General

Merchandise and Container (TEUs). In Figures 3 through 12 show the Dissimilarity index and the Gini Coefficient of this type of goods.
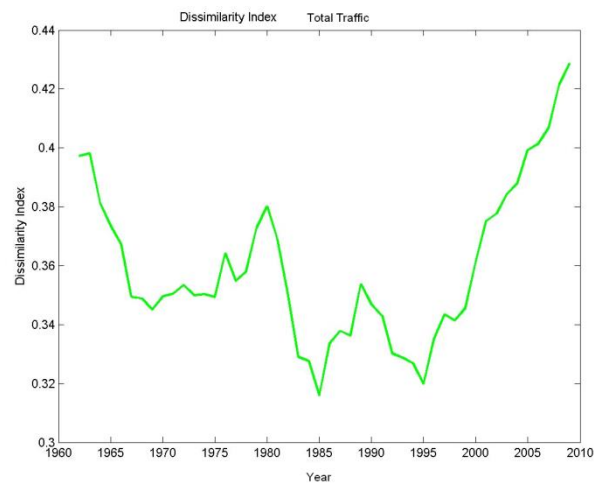

Figure 3.   Dissimilarity Index. Liquid Bulk Cargo. Total Traffic
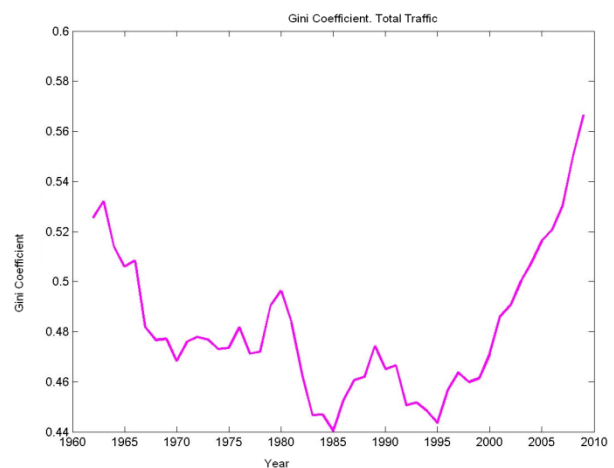

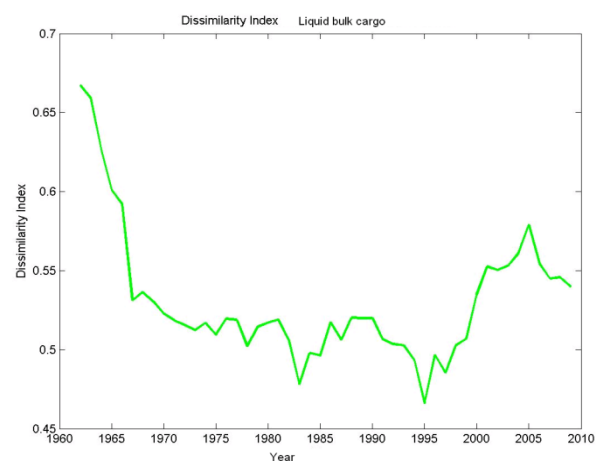Figure 4.   Gini Coefficient. Total Traffic

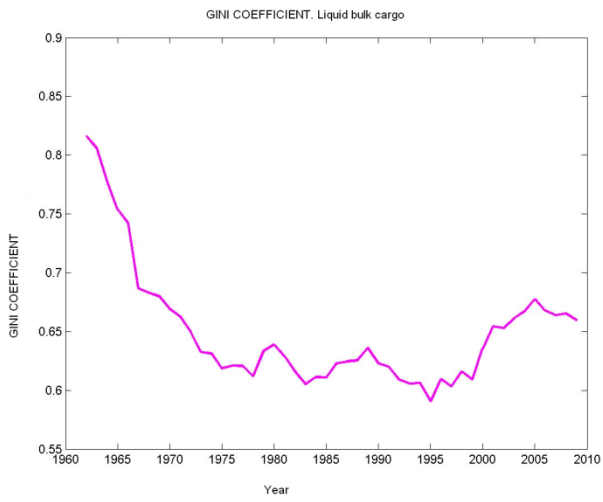Figure 5.   Dissimilarity Index. Liquid Bulk Cargo.

Figure 6.   Gini Coefficient. Liquid Bulk Cargo
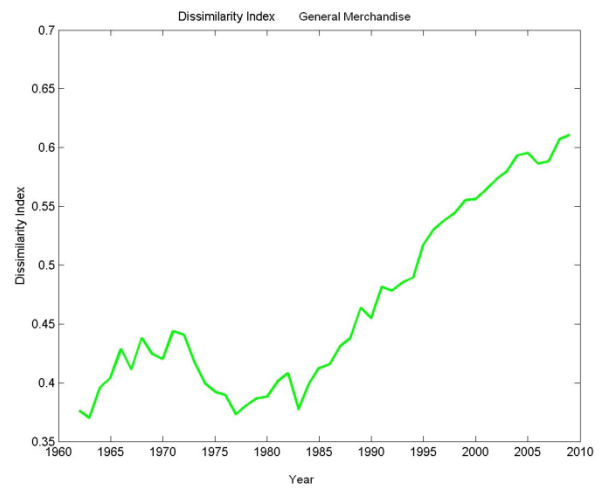

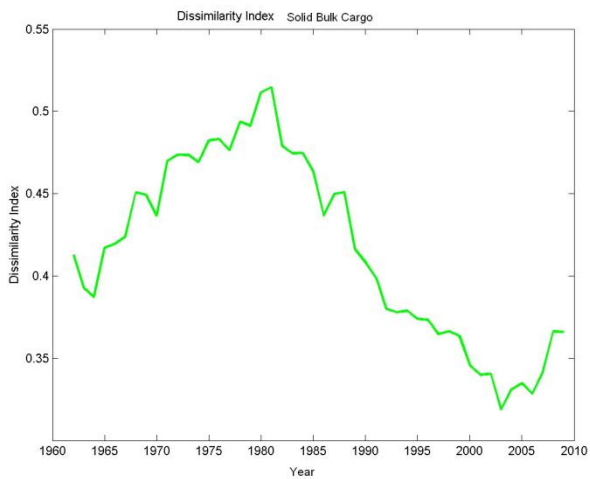Figure 9.   Dissimilarity Index. General Merchandise

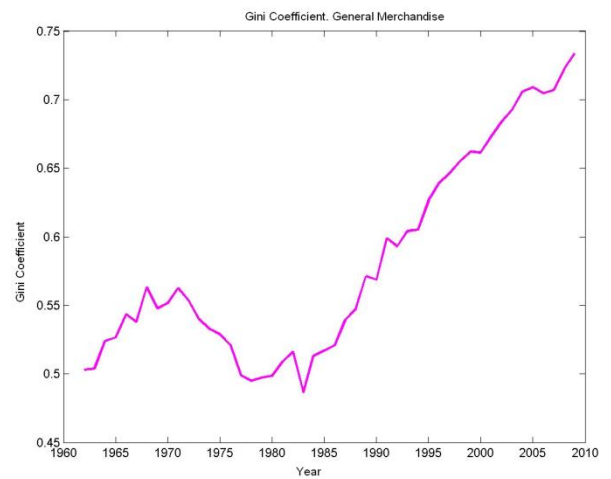Figure 7.   Dissimilarity Index. Solid Bulk Cargo

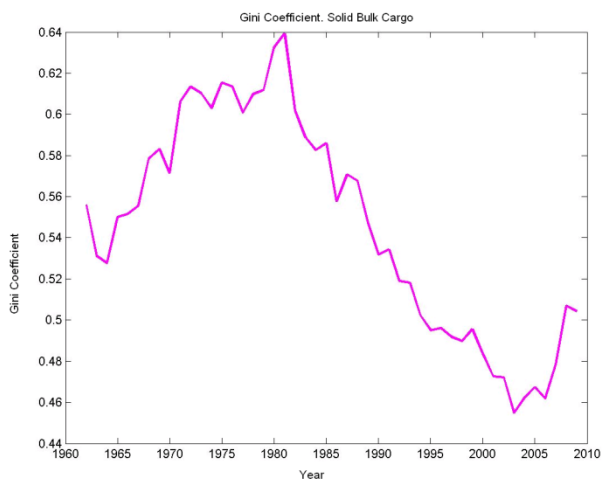Figure 10. Gini Coefficient. General Merchandise

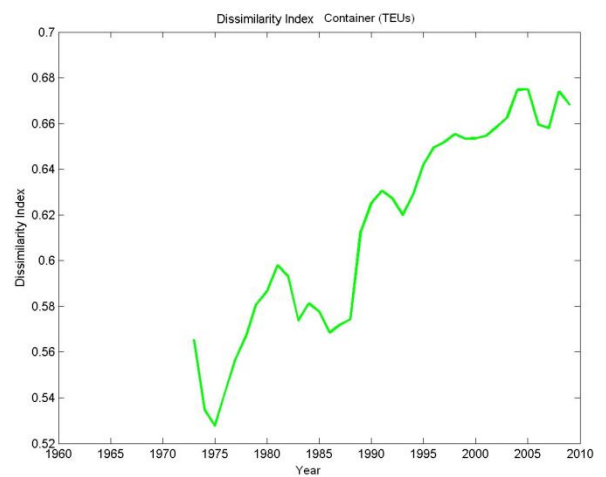Figure 8.   Gini Coefficient. Solid Bulk Cargo

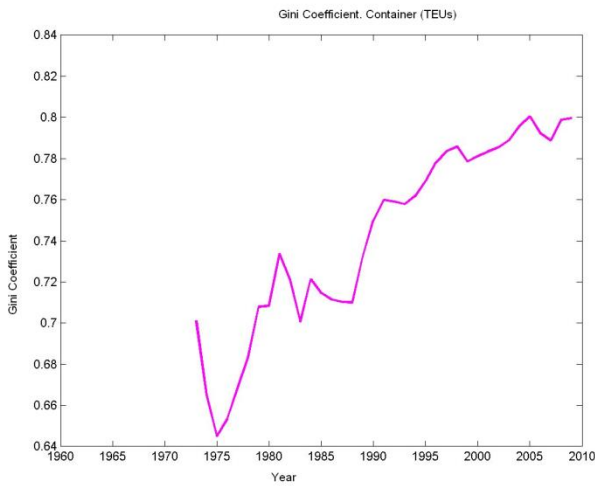Figure 11. Dissimilarity Index. Container (TEUs)

Figure 12. Gini Coefficient. Container (TEUs)

For each year of study and each type of merchandise is estimated Lorenz curve, below are some examples in the figures 13, 14.
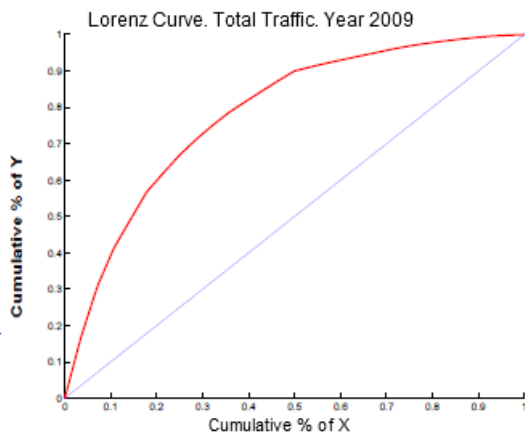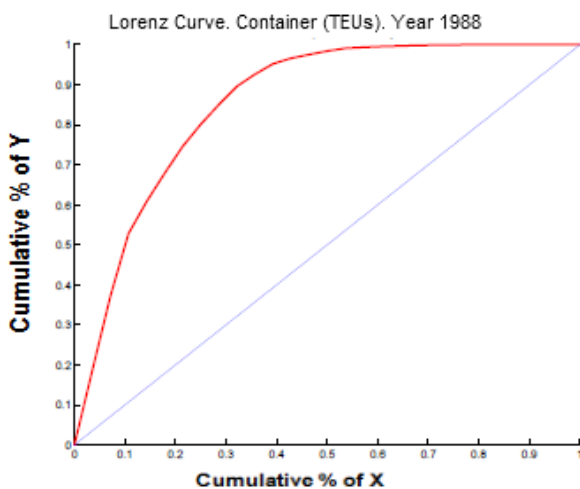

Figure 13. Lorenz Curve. Total Traffic. Year 2009

Figure 14. Lorenz Curve. Container (TEUs). Year 1988

IV. CONCLUSIONS

Total Traffic:

Both DI and GC exhibit pronounced peaks (closer to 1 greater tendency to inequality in the distribution, merchandise is concentrated in a few ports) in the years 1963, 1975, 1980, 1990 and 2009, which coincides with the first and second crisis oil (1975 and 1978), and after the current crisis is a tendency to concentrate on efficient ports merchandise. The valleys appear in 1985 and 1995 after the following industrial restructuring (1983) where it is close to perfect equality and after the crisis to Latin America in 1994.

Liquid Bulk Cargo:

Both DI and GC have the same type, but the GC with more pronounced peaks and valleys. The peaks are more pronounced in 1963 and 2005, after the invasion of Iraq, and the valleys in 1983 and 1995 after the Iran-Iraq war (1981) after the crisis with Latin-America in 1994.

Solid Bulk Cargo:

Both DI and GC have the same type, but the GC with more pronounced peaks and valleys. The peaks are more pronounced in 1979 and 2009 and troughs in 1965 and 2003. The peaks appear after the second oil crisis in 1978 and after the onset of the current crisis. And the valleys after the attacks of 11-S in Madrid and the invasion of Iraq.

General Merchandise:

Both DI and GC exhibit pronounced peaks (closer than 1 greater tendency to inequality in the distribution, merchandise is concentrated in a few ports) 1965 to 1975, before the oil crisis and 2009 after the start of the current crisis . The valleys are presented between 1977 and 1979 between the first and second oil crisis.

Container (TEU):

Both DI and GC have the same type, but the GC with more pronounced peaks and valleys. The peaks seen in 1980 and 2005 after the oil crisis and the invasion of Iraq and the valley in 1975 and 1983 after the Iran-Iraq war.

We conclude that for the Spanish port system, the oil crisis tendencies have resulted in concentrating the goods in little more efficient ports and went after the invasion of Iraq. By contrast, the Latin American crisis and the Iran-Iraq War was a tendency for the distribution of goods in all ports of the Spanish port system.

This indicates that as long as the crisis scenario the tendency will be to present both GC, DI as pronounced peaks, closer to 1 indicating greater tendency to inequality in the distribution, and is concentrated in a few ports merchandise.

Regarding the study of the Lorenz curve shows that the total traffic tends to perfection inequaly line, more prominent in the case of container traffic.

REFERENCES

[1] Quijada-Alarcon, N. Gonz alez Cancelas, A. Camarero Orive y F. Soler Flores, "Road network analysis using decision trees algorithm: A case of

study of Panama", Proceedings in Advanced Research in Scientic Areas. The 1st Virtual International Conference, 2012.

[2] UNCTAD. "Desarrollo Portuario. Manual de Planificación Para Países en Desarrollo". Naciones Unidas, Nueva York, 1984

[3] Soler, R. "Índices Portuarios Españoles". Revista de Obras Públicas, Febrero 1979.

[4] Fourgeaud, P., "Measuring Port Performance", World Bank TWUTD, mimeo, 2000

[5] Drewry, "World Container Terminals", Drewry Shipping Consultants, 1997.

[6] Drewry, "Global Port Congestion: No Quick Fix, Drewry Shipping Consultants", February 2005.

[7] Schreuder, M "Application of Approximate Performance Indicators for Master Planning of Large Ports". Port Technology International. Nº26, 2005, pp 19-22

[8] BTRE.– Department of Transport and Regional Services, Bureau of Transport and Regional Economics, Australia, Waterline issue nº 41, December 2006

[9] Agerschou, Hans, "Facilities Requirements. In Planning and Design of Ports and Marine Terminals", 2nd Ed. (H. Agerschou, eds), Thomas Telford Ltd., Londres, 2004, pp. 5-20.

[10] Dragovic, B., Zrnic, D. y Radmilovic, Z. "Ports & Container Terminals Modeling". Research Monograph. Faculty of Transport and Traffic Engineering, University of Belgrade. 2006.

[11] Nicolaou, S.N. "Berth Planning by Evaluation of Congestion and Cost", Journal of the Waterways and Harbors Division, Proceedings of the American Society of Civil Engineers, 1967, n 93

[12] Nicolaou, S.N. "Berth Planning by Evaluation of Congestion and Cost - Closure". Journal of the Waterways and Harbors Division, Proceedings of the American Society of Civil Engineers, 95 (WW3), 1969, pp. 419-425

[13] Plumlee, C.H. "Optimum Size Seaport", Journal of the Waterways and Harbors Division, Proceedings of the American Society of Civil Engineers, vol 92, 1966, pp. 1-24.

[14] UNCTAD. "Desarrollo de los Puertos. Mejoramiento de las Operaciones Portuarias e Instalaciones Conexas". Naciones Unidas, Nueva York, 1969.

[15] Maloni, M. J.; Jackson, E. C., "North American Container Port Capacity: A [Literature Review", Transportation Journal 44 (2), 2005, pp.16-36

[16] Fan, H.S.L.; Cao, J.M., "Sea Space Capacity and Operation Strategy Analysis System", Transportation Planning and Technology, 2000, pp. 24, 49-63

[17] Camarero, A., M. González, "Optimización y estudio de la capacidad de las terminales portuarias mediante modelos de simulación y explotación. Determinación de los niveles de servicio." UPM, UPV, Cenit, Valencia Port, Cedex, España, 2009

[18] González, MN. "Metodología para la determinación de parámetros de diseño de terminales portuarias de contenedores a partir de datos de tráfico marítimo", Thesis (UPM), 2007

[19] Warren S. McCulloch, Walter H.Pitts, "A logical calculus of the ideas inmanet in nervous activity"., Bulletin os Mathematical Biophsica, Vol5, 1943, pp. 115-133.

[20] Hopfield, John, "Neural networks and physical systems with emergent collective computational abilities." Proceedings of the National Academy of Sciences of the USA, vol. 79, no. 8, April 1982, pp. 2554-2558

[21] Moscoso, J.A., Ruiz, J.J., Cerbán, MM., "Técnicas de predicción en el tráfico Ro-Ro en el nodo logístico del Estrecho de Gibraltar", IAME 2011 conference

[22] Veerachai Gosasang ,Watcharavee Chandraprakaikul, Supaporn Kiattisin. "An Application of Neural Networks for Forecasting Container Throughput at Bangkok Port". Proceedings of the World Congress on Engineering 2010 Vol I WCE 2010, June 30 - July 2, 2010, London, U.K.ISBN: 978-988-17012-9-9

[23] Gosasang, Veerachai, Watcharavee Chandraprakaikul, And Supaporn Kiattisin. "A Comparison of Traditional and Neural Networks Forecasting Techniques for Container Throughput at Bangkok Port." The Asian Journal of Shipping and Logistics. Volume 27.3, 2011, pp. 463-482.

[24] Karlaftis M.G., Vlahogianni E.I. "Statistical methods versus neural networks in transportatioin research: Differences, similarities and some insights". Transportation Research Part C 19, 2011, pp. 387-399

[25] Vlahogianni E, Karlaftis M, Golias J, "Optimized and meta-optimized neural networks forshort-term traffic flow prediction: A genetic approach", Transportation Research Part C 13, 2005, pp. 211–234

[26] Dougherty M. S. and Cobbett M. "Short term inter-urban traffic forecasts using neural networks". Proc. 2nd DRIVE-II Workshop on Short-Term Forecasting, Delft.The Netherlands. 1994

[27] Dougherty M. S., Kirby H. R. and Boyle R. D. "Using neural networks to recognise predict and model traffic". Artificial Intelligence Applications to Traffic Engineering (Bielli, Ambrosino and Boero, Eds). VSP, Utrecht. 1994.

[28] Clark S. D. Dougherty M. S. and Kirby H. R. "The use of neural network and time series modes for short term forecasting: a comparative study". Proc. PTRC Summer Meeting, Manchester., 1993.

[29] Tsai, T.-H., Lee, C.-K., Wei, C.-H., "Neural network based temporal feature models for short-term railway passenger demand forecasting". Expert Systems with Applications 36, 2009, pp.3728–3736.