

Clasificación de los puertos españoles mediante análisis *cluster*

Classification of Spanish ports using cluster analysis

A. Camarero (*), A. Camarero (*), M. M. Cerbán (**), I. J. Turias (**), N. González-Cancelas (*)

RESUMEN

El sistema portuario español es sumamente complejo y admite el estudio desde numerosos puntos de vista. En este artículo se estudian los puertos según su actividad y sus características externas para la clasificación en agrupaciones. Para ello se han utilizado indicadores que reflejan la actividad portuaria y se han aplicado sobre las 28 Autoridades Portuarias españolas. Con estos indicadores se ha aplicado una metodología específica para el análisis de conglomerados (*cluster*) para averiguar cuáles son los agrupamientos que se producen. El análisis *cluster* se complementa con otros análisis estadísticos: análisis multivariante y componentes principales, para conocer qué indicadores son los más relevantes en las agrupaciones y cómo se comportan. Los resultados finales obtenidos muestran que este tipo de estudios estadísticos son apropiados para realizarse en el entorno portuario y que los agrupamientos reflejan correctamente la realidad portuaria.

Palabras clave: Puerto, metodología, clasificación, indicadores, conglomerados, *cluster*.

ABSTRACT

The Spanish port system is extremely complex and admits the study from many points of view. In this article the ports are studied from the point of view of classification in clusters according to their external characteristics. For this purpose, indicators have been used that reflect the port activity and have been applied on the 28 Spanish Port Authorities. With these indicators, a specific methodology has been applied through the analysis of clusters (cluster) to find out which clusters are produced. The cluster analysis is complemented by other analyzes (main components, multivariate analysis and individual indicators) to know which indicators are the most relevant in clusters and how they behave. The final results obtained show that this type of statistical studies are appropriate to be carried out in the port environment and that the groupings correctly reflect the port reality.

Keywords: Port, methodology, classification, indicators, cluster.

(*) Universidad Politécnica de Madrid (España).

(**) Universidad de Cádiz (España).

Persona de contacto/Corresponding author: alfonso.c.orive@gmail.com (A. Camarero).

ORCID: <https://orcid.org/0000-0002-8274-8383> (A. Camarero); <https://orcid.org/0000-0002-0614-8714> (A. Camarero);

<https://orcid.org/0000-0003-4674-0453> (M. M. Cerbán); <https://orcid.org/0000-0003-4627-0252> (I. J. Turias);

<https://orcid.org/0000-0001-7167-1563> (N. González-Cancelas).

Cómo citar este artículo/Citation: Camarero, A.; Camarero, A.; Cerbán, M. M.; Turias, I. J.; González-Cancelas, N. (2019). Clasificación de los puertos españoles mediante análisis *cluster*. *Informes de la Construcción*, 71(554): e296. <https://doi.org/10.3989/ic.61806>

Copyright: © 2019 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

Recibido/Received: 15/11/2017
Aceptado/Accepted: 12/06/2018
Publicado on-line/Published on-line: 10/07/2019

1. INTRODUCCIÓN

En el entorno portuario se vienen realizando numerosos estudios basados en el tráfico marítimo y las operaciones portuarias. Es común en estos estudios centrarse en criterios muy variados, como pueden ser la capacidad de la línea de atraque, la distribución de los muelles, la eficiencia en la operativa de carga/descarga, la seguridad, etc. El propósito es conocer las ventajas e inconvenientes en un escenario concreto, con el objetivo de mejorar la operatividad y el rendimiento portuarios, ya que estos informes sirven para diseñar estrategias de desarrollo. Sin embargo, también se puede considerar el entorno portuario adecuado para realizar otros estudios de mercado en base a conceptos que abarcan desde la eficiencia y la competitividad, hasta el *marketing*. En este artículo se tratan los puertos, no de manera individual, sino conjunta, y lo que se pretende es compararlos entre sí en base a algunos criterios definidos de antemano, mediante el uso de indicadores que puedan caracterizar un puerto.

Si se considera que la clasificación es uno de los objetivos fundamentales de la ciencia, el análisis *cluster* proporciona los medios técnicos para realizarla. Es una herramienta utilizada para realizar estudios de clasificación o de aproximación general a un entorno, donde los grupos de muestreo pueden ser numerosos o pequeños, y elegidos con intencionalidad o al azar. Se trata de un tipo de clasificación de objetos exploratoria cuya estructura resultante permite simplificar el conjunto de observaciones. Con esta herramienta también es posible identificar relaciones que estaban ocultas y los agrupamientos mostrarán las similitudes y diferencias que presenten los elementos estudiados.

Para lograrlo, es necesario definir los indicadores que reflejen la actividad portuaria de la manera más completa posible. Esta investigación se ha servido de un estudio previo con 35 expertos portuarios sobre la definición y caracterización de los puertos mediante variables, donde se han utilizado paneles Delphi y dos rondas de resultados. Aún así, entre todos los indicadores seleccionados, unos pocos serán los más determinantes a la hora de realizar los agrupamientos, los llamados indicadores principales, y son, evidentemente, los más diferenciadores, los que otorgan mayor carácter. Para encontrarlos, el análisis de componentes principales juega un papel destacado, complementado con otros análisis estadísticos multivariante.

En este trabajo se ha llevado a la práctica una metodología específicamente diseñada, que se aplica para realizar un análisis de conglomerados sobre los puertos españoles basándose en indicadores físicos y de explotación.

2. REVISIÓN DE LA LITERATURA

En el entorno económico, social y político actual, la sociedad exige una mayor variedad de los resultados del sector de la logística pública, como la eficiencia, la eficacia de los recursos gestionados, mayor transparencia y rendimiento comercial. Todos ellos son una contraparte indispensable para su reconocimiento y apoyo. En caso de planificación y gestión de puertos, se incluyen muchas variables.

De la misma manera que se hace en las ciudades (1), es posible realizar estudios de investigación sobre numerosos te-

mas a través de indicadores en el entorno portuario, aunque los trabajos de investigación con indicadores en los puertos son mucho menos variados, pero no menos interesantes (2). Sin embargo, con ellos es posible analizar la eficiencia y la productividad del sector portuario (3), por lo que merecen ser tenidos en mucha consideración. Especialmente, la competencia portuaria en el tráfico de contenedores ha propiciado que se aborden múltiples trabajos y análisis de comparación entre terminales y operadores (4). También toman relevancia los estudios con indicadores para medir el impacto ambiental en los puertos, principalmente los más industrializados (5), así como para establecer herramientas que permitan medir la sostenibilidad en obras de ingeniería civil según sea su impacto sobre la economía, la sociedad o la economía (6).

Debido a la proliferación de datos y el acceso público a los mismos, ha sido posible que la mayoría de estas investigaciones analicen enormes conjuntos de datos, utilizando para ello las técnicas estadísticas multivariante. Dentro de ellas, el análisis de conglomerados trata de obtener clasificaciones de grupos, y un conglomerado (*cluster*) es un conjunto de objetos que contiene datos similares y distintos a los de otro conglomerado (homogeneidad interna y heterogeneidad externa). De manera, que resulta muy eficaz para encontrar patrones desde varios puntos de vista, según la relación simultánea en las dimensiones que se han definido.

Las descripciones más tempranas de la técnica conocida como Análisis de Componentes Principales (ACP o PCA en inglés) fueron dadas por Pearson en 1901 (7) y Hotelling a partir de 1930 (8). Al igual que los trabajos de Fisher (9), marcaron el inicio de los estudios de agrupamientos, que se han extendido por todos los ámbitos científicos, especialmente con el desarrollo de las computadoras y el manejo de enormes bases de datos, y hoy se encuentran en prácticamente todas las disciplinas científicas (10).

Desde hace pocos años el estudio de los agrupamientos se lleva realizando también en el sector logístico (11). En Suiza, se presentó un sistema de clasificación de las estaciones de ferrocarril para el transporte estratégico y la planificación del uso del suelo (12) con un análisis de conglomerados de las 1.700 estaciones de ferrocarril suizas que dependen exclusivamente de los factores contextuales, donde las clases resultantes variaban principalmente en la densidad y el uso. En el sector portuario también se han realizado estudios mediante el análisis *cluster* (13). Sobre la competitividad y la competencia portuaria, se han realizado estudios entre puertos internacionales de distintos continentes (14), y en España también se realizaron trabajos sobre los factores determinantes de la competitividad y competencia portuaria más relevantes en los puertos españoles a través de un grupo de indicadores sobre infraestructuras y servicios portuarios (15). Otros estudios relacionados con la competitividad resaltan los factores que influyen en la selección de un puerto en el sistema español (16), que son los que reflejan la capacidad de captación de tráfico. Mientras que otros trabajos (17) se han centrado en el modelo de eficiencia de las terminales de contenedores.

El desarrollo sostenible se está aplicando por las autoridades de transporte, de la misma manera que se ha realizado en otros sectores de actividad e industrias en todo el mundo. Así, el transporte debe incluir variables del medio am-

biente y responsabilidad social en la gestión estratégica de las empresas y en las economías globales (18). El desarrollo sostenible busca un equilibrio en las dimensiones económicas, sociales, ambientales e institucionales, y también se manifiesta en la operativa portuaria (19). En este contexto, el transporte marítimo requiere una atención especial, porque las operaciones de envío contribuyen al crecimiento de las actividades comerciales internacionales (20), ya que transporta más del 80% de las transacciones en el extranjero (TON-Km) (21).

3. METODOLOGÍA

La metodología que se ha desarrollado en esta investigación se apoya en varias técnicas para abordar la complejidad del entorno portuario, que comprende: análisis del entorno, caracterización de los puertos, búsqueda de fuentes de datos, estudio de los indicadores, toma de datos, aplicación de los análisis estadísticos y revisión de los resultados obtenidos. La Figura 1 muestra el esquema que se ha seguido en la metodología, dividido en tres etapas.

En primer lugar, es preciso acotar el escenario de trabajo y delimitar las condiciones que deben cumplir los puertos del estudio, ya que éstas pueden estar referidas al tamaño, la actividad, la ubicación geográfica, el tipo de tráfico de mercancías, etc. Los datos que se manejan deben cumplir ciertas garantías de calidad para poder ser manejados en los análisis: filtrado, eliminación de anomalías o de registros duplicados (13). El análisis de fuentes de datos incluye, no sólo la forma de recolección, estudio y procesamiento de datos, sino también afecta a la selección de los puertos examinados, puesto que uno de los más graves problemas para este tipo de trabajos de investigación es la disponibilidad de datos, sin los cuales no se podría realizar la posterior clasificación. Además, estos datos deben cumplir ciertas características como ser: objetivos, medibles, públicos, etc., y se prefieren a los que son imprecisos, redundantes o poco homogeneizados con sus valores.

Para caracterizar un puerto bajo un determinado punto de vista es necesario definir una serie de indicadores que reflejen todo lo que se quiere medir. El proceso metodológico de elaboración de indicadores es una tarea compleja. Los indicadores portuarios resumen las características de un puerto y hacen referencia al tipo de explotación a la que se dedica el puerto, e incluyen factores físicos, de estructuras, maquinaria, dimensionamiento, etc., así como de su funcionamiento y operativa. En definitiva, hay que llegar a definir un puerto según una serie de variables y que éstas sean las mínimas posibles. Una vez definidos los indicadores se procede a su valoración a través de la recogida de datos disponibles. Todos los indicadores deben tomar un valor concreto para cada puerto considerado.

Para el análisis estadístico se ha escogido el análisis de conglomerados *cluster* que realiza agrupamientos según la afinidad que tengan los puertos en base a la similitud de sus características, obteniendo así clasificaciones (22). Con este análisis se identifican las relaciones multivariante que difícilmente son accesibles a partir del análisis de las unidades espaciales individuales. Se establece una matriz de distancias entre los individuos del fichero de datos antes de la clasificación: $(n \times k)$ de n objetos (puertos) y k variables (indicadores portuarios), donde cada objeto queda definido por el valor de sus variables.

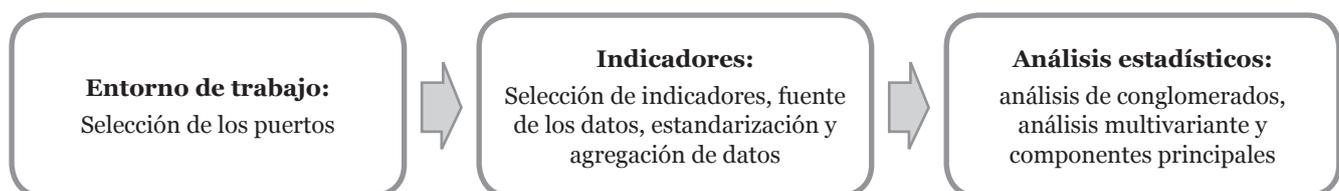
Para formar los conglomerados, el procedimiento comienza con cada observación en grupos separados, y los individuos que guardan menor distancia serán los primeros que se unan en un mismo grupo, sabiendo que los conglomerados son grupos de observaciones con características similares. El grupo formado por estos dos individuos se modificará más adelante con nuevos elementos. Si dos objetos P1 y P2 se han agrupado, la distancia de grupos con otro objeto P3 puede calcularse en función de las distancias de la siguiente forma:

$$d(P_3, P_1 + P_2) = \delta_1 d(P_3, P_1) + \delta_2 d(P_3, P_2) + \delta_3 d(P_2, P_3) + \delta_4 |d(P_3, P_1) - d(P_3, P_2)| \quad [1]$$

Donde δ_j son constantes de ponderación.

Después de recalcular la distancia entre grupos, se combinan los dos grupos ahora más cercanos. Este proceso se repite hasta que queda solamente el número de grupos que se ha predeterminado. De manera, que va aumentando la variabilidad de los conglomerados conforme se van añadiendo elementos. En la primera etapa estarían todos los objetos como grupos aislados y en la última etapa quedará un grupo que engloba a todos los objetos. El análisis *cluster* es muy sensible a la presencia de objetos atípicos, que inevitablemente serán los últimos en agrupar. El algoritmo utilizado para la obtención de los conglomerados está basado en procedimientos jerárquicos, prefiriendo el vecino más lejano porque implica un encadenamiento completo y es útil para detectar valores atípicos.

Los resultados del análisis de conglomerados son varios: la tabla de centroides muestra el valor promedio para cada variable en cada conglomerado final; el historial de conglomeración muestra los niveles de fusión (coeficientes) al que se van uniendo los individuos en los conglomerados en cada etapa; y el dendrograma, que es la representación gráfica más característica del análisis *cluster*. El dendrograma tiene forma de árbol jerárquico que indica secuencialmente las uniones de las unidades espaciales, permitiendo definir el punto de corte de mayor aptitud y con ello la regionalización. Se ali-



Fuente: Elaboración propia

Figura 1. Etapas de la clasificación clustering de las Autoridades Portuarias.

nean en un eje todos los objetos y en el otro eje la distancia que les separa, lo que implica un nivel de jerarquía.

En la metodología que se ha seguido para este trabajo de investigación, se realizan una serie de análisis complementarios para estudiar mejor los agrupamientos formados. Con análisis multivariante se pueden calcular estadísticos de resumen para cada variable para revisar si existen desviaciones de la distribución normal y la relación lineal entre las variables. El gráfico de rayos es una técnica de visualización muy útil para identificar diferencias y similitudes entre casos observados cuando el número de dimensiones es demasiado grande para emplear un diagrama de dispersión estándar. Aplicado al análisis de conglomerados, las figuras más parecidas son las que antes se agrupan entre sí.

El análisis de componentes principales (10) es una técnica complementaria al análisis de conglomerados para reducir el conjunto de variables que determinan los grupos, con el fin de descubrir qué indicadores son más significativos y discriminatorios. De esta manera, los datos quedarán mejor representados en términos de mínimos cuadrados. El procedimiento de componentes principales está diseñado para extraer k componentes principales de un conjunto de p variables cuantitativas X . Los componentes principales están definidos como el conjunto de combinaciones lineales ortogonales de X con la varianza más grande. Cuando las variables están altamente correlacionadas, los primeros componentes principales pueden ser suficientes para describir la mayoría de la variabilidad presente. La meta de un análisis de componentes principales es construir k combinaciones lineales de las p variables X que tengan la mayor varianza. Las combinaciones lineales toman la forma:

$$\begin{aligned} Y_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \\ Y_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \\ Y_k &= a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kp}x_p \end{aligned} \quad [2]$$

El primer componente principal es aquella combinación lineal con la mayor varianza, sujeto a la restricción de que el vector de coeficientes tiene longitud unitaria, es decir:

$$\sum_{i=1}^p a_{1i}^2 = 1 \quad [3]$$

Si la matriz de covarianza de X es igual a Σ entonces la varianza de Y_1 es:

$$\text{Var}(Y_1) = a_1' \Sigma a_1 \quad [4]$$

El segundo componente principal es aquella combinación lineal con la segunda mayor varianza, sujeta a la misma restricción de longitud, y además no correlacionada al primer componente principal. Los demás componentes son aquellos con la mayor varianza posible, además de no estar correlacionados con ningún otro de los componentes. Bajo este modelo, los coeficientes a corresponden a los autovectores de Σ , mientras que las varianzas de las Y 's son iguales a los autovalores

$$\text{Var}(Y_j) = \lambda_j \quad [5]$$

La varianza poblacional es igual a la suma de los autovalores.

La tabla de pesos muestra los valores estimados de los coeficientes para cada componente extraído. Si los pesos en una columna son todos aproximadamente iguales, esto implica que el componente es básicamente un promedio de todas las variables de entrada. Si el componente está ponderado más en una dirección, positiva o negativa, probablemente esas variables marcan diferencia entre distintos tipos de elementos.

Finalmente, en la metodología propuesta para este estudio, después de extraer las componentes principales de las agrupaciones, se identifican también las variables (indicadores) más significativas de las que se han utilizado para caracterizar un puerto.

4. RESULTADOS

La presente investigación se ha centrado en el sistema portuario de titularidad estatal en España. El sistema portuario español se compone de 28 Autoridades Portuarias que gestionan 46 puertos de interés general en un régimen avanzado de autonomía de gestión. Se han recogido los datos directamente de las Memorias de de las Autoridades Portuarias y de los Anuarios Estadísticos de Puertos del Estado (23). Además, las terminales del sistema portuario español son muy heterogéneas debido a la diferente tipología de sus tráficos, la diversidad de sus equipos de manipulación de mercancías y la variedad en su gestión y explotación (24) y (25).

Los indicadores que se han seleccionado para esta investigación son los más comúnmente usados en los estudios del sistema portuario y han sido consensuados por expertos para eliminar otros indicadores que resulten menos relevantes. El grupo de indicadores corresponde a factores genéricos que tienen que ver con tráfico y el tipo de explotación portuaria, con las características físicas de los puertos y otro referido a las inversiones económicas. En cuanto a los indicadores seleccionados para caracterizar a los puertos, se han tomado los de tipo cuantitativo. No se usa en esta metodología la ponderación en la importancia de los indicadores, es decir, se considera que todos los indicadores tienen la misma influencia, así que el análisis *cluster* solamente compara la semejanza en la composición relativa de estas características. No se han hecho aproximaciones ni simplificaciones. Tampoco existen valores nulos. En total, se han seleccionado 23 indicadores, que se muestran en la Tabla 1 junto con su definición y las unidades en las que se miden.

Una vez preparada la base de datos que se va a manejar se pasa a los análisis estadísticos. Para el análisis de conglomerados se debe establecer el número de patrones, la estandarización, la definición de proximidad, el tipo de *cluster* y la representación de los datos. El proceso de agrupamiento se inicia con los dos elementos más próximos y se repite hasta que queda 1 solo grupo a partir de las 28 observaciones proporcionadas.

La Tabla 2 muestra la matriz de distancias. Así se conoce qué observaciones se combinaron en cada etapa del proceso. Por ejemplo, en la primera etapa, se combinaron la observación 19 (Melilla) con la observación 20 (Motril), lo que significa que estos son los elementos más parecidos. La distancia entre los grupos, una vez combinado, fue 0,467986. También se muestra que la siguiente etapa en la que este grupo combinado se combinó con otro conglomerado fue la etapa 4.

Tabla 1. Selección de indicadores portuarios.

Identificador	Nombre	Definición	Unidad
IG_1	Superficie de flotación Zona1	Zonas portuarias formadas por las superficies de agua interior por donde se desplazan los buques en el recinto portuario	Hectáreas
IG_2	Superficie de flotación Zona2	Zonas portuarias formadas por las superficies de agua exterior por donde se desplazan los buques en el recinto portuario	Hectáreas
IG_3	Calado superior a 4 metros	Longitud total de muelles con un calado superior a los 4 metros	m
IG_4	Muelles	Longitud total de los muelles	m
IG_5	Superficie terrestre de almacenamiento	Superficie total en el puerto dedicada al almacenamiento de mercancías y áreas de depósito	m ²
IG_6	Almacenes frigoríficos	Volumen del espacio dedicado en el puerto al almacenaje refrigerado	m ³
IG_7	Grúas	Número de grúas de todos los tipos utilizadas para la carga y descarga de buques	n
IG_8	Remolcadores	Número de remolcadores utilizados para ayudar a maniobrar a los buques	n
IG_9	Ayudas a la navegación	Sistemas de ayuda a la navegación marítima	n
IG_10	Productos petrolíficos	Tráfico de graneles líquidos sólo de productos petrolíferos que han circulado por el puerto en el año	t
IG_11	Otros productos líquidos	Tráfico de graneles líquidos de productos no petrolíferos que han circulado por el puerto en el año	t
IG_12	Graneles sólidos	Tráfico de graneles sólidos que han circulado por el puerto en el año	t
IG_13	Mercancía general convencional	Tráfico de mercancía general convencional que ha circulado por el puerto en el año	t
IG_14	Mercancía general contenedores	Tráfico de mercancía general en contenedores que ha circulado por el puerto en el año	t
IG_15	Pesca	Cantidad total de pesca capturada en el año	t
IG_16	Avituallamiento combustible	Cantidad de combustible suministrado a los buques en el puerto	t
IG_17	Avituallamiento restante	Otros avituallamientos suministrados a los buques en el puerto	t
IG_18	Contenedores	Número total de contenedores que han pasado por el puerto en el año	TEU
IG_19	Buques	Número total de buques que ha pasado por el puerto en el año	n
IG_20	Pasajeros de cruceros	Número total de pasajeros de cruceros que ha pasado por el puerto en el año	n
IG_21	Pasajeros	Número total de pasajeros que ha pasado por el puerto en el año	n
IG_22	Vehículos	Número total de vehículos en régimen de pasaje que ha pasado por el puerto en el año	n
IG_23	Inversiones	Inversiones totales destinadas en el puerto	Miles euros

Fuente: Elaboración propia.

Tabla 2. Historial de conglomeración y de distancias del *clustering*.

Etapa	Conglomerado 1 Combinado	Conglomerado 2 Combinado	Distancia	Etapa Previa Conglomerado 1	Etapa Previa Conglomerado 2	Etapa Siguiente
1	19	20	0,467986	0	0	4
2	4	21	1,05265	0	0	5
3	2	17	1,28088	0	0	6
4	19	28	2,13423	1	0	7
5	3	4	2,73068	0	2	6
6	2	3	3,36635	3	5	7
7	2	19	3,56563	6	4	8
8	2	18	4,22103	7	0	10
9	10	25	5,82799	0	0	15
10	2	12	6,74082	8	0	16
11	11	24	7,95292	0	0	13
12	6	23	10,4261	0	0	13
13	6	11	12,7064	12	11	14
14	1	6	18,0044	0	13	16
15	9	10	22,0035	0	9	18
16	1	2	22,9639	14	10	22
17	13	14	26,0233	0	0	18
18	9	13	31,2626	15	17	20
19	8	16	32,2445	0	0	21
20	9	15	48,4866	18	0	25
21	8	26	51,3625	19	0	24
22	1	27	58,2031	16	0	25
23	7	22	61,5911	0	0	24
24	7	8	85,8283	23	21	26
25	1	9	97,5791	22	20	27
26	5	7	116,875	0	24	27
27	1	5	168,678	25	26	0

Fuente: Elaboración propia.

El dendrograma con los agrupamientos según la distancia aparece en la Figura 2. En donde se conoce qué observaciones se combinaron en cada etapa del proceso. Se puede comprobar que las primeras agrupaciones se realizan entre elementos muy semejantes y con poca distancia entre ellos.

Analizando estos resultados, se pueden establecer 3 grupos finales diferenciados. El primer grupo está formado por la ma-

yoría de los puertos (22 AP), es decir, lo integra el 78'57%, el segundo grupo (1 AP) presenta Bahía de Algeciras como único elemento, es decir, el 3'57% y el tercer grupo es poco homogéneo y contiene pocas Autoridades Portuarias (5 AP): Baleares, Barcelona, Las Palmas, Santa Cruz de Tenerife y Valencia, es decir, lo integra el 17'86%. En la Tabla 3 aparecen los centroides de los indicadores para estos tres grupos que se pueden comparar con con los valores individuales de cada puerto.

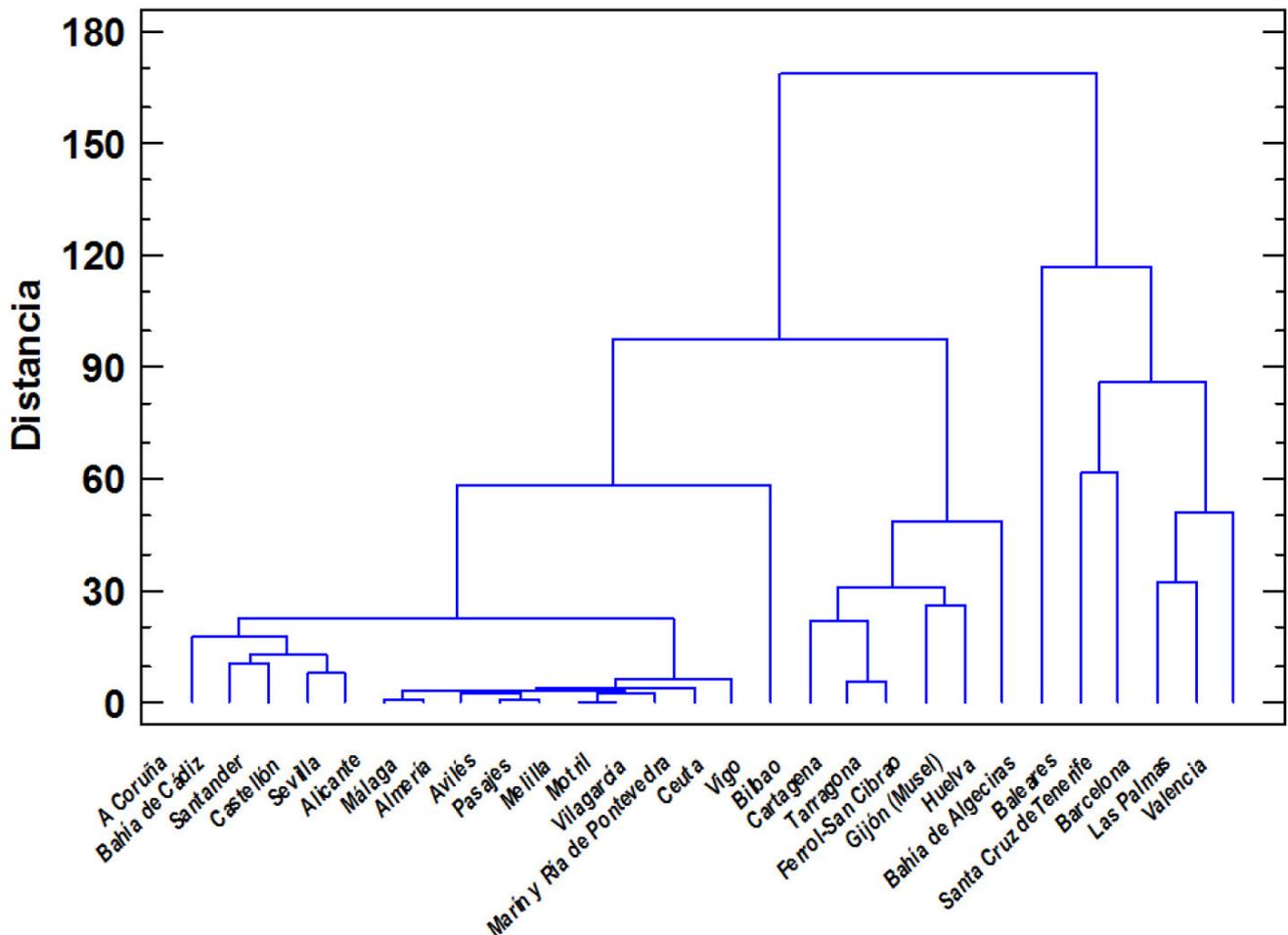
Tabla 3. Tabla de centroides con el valor de los indicadores IG para tres grupos finales.

Conglomerado	IG_1	IG_2	IG_3	IG_4	IG_5	IG_6	IG_7	IG_8	IG_9
1	599,86	5142,32	8334,86	8309,86	2,94E6	69643,70	15,81	5,36	56,40
2	1010,00	5251,00	21444,00	21444,00	5,95E6	36123,00	144,00	13,00	82,00
3	1017,60	22299,00	20329,80	20329,80	5,52E6	75037,80	29,60	8,20	109,80

Conglomerado	IG_10	IG_11	IG_12	IG_13	IG_14	IG_15	IG_16	IG_17
1	3,88E6	907976,00	3,55E6	926615,00	900222,00	9985,68	49230,10	36295,30
2	2,40E7	1,16E6	1,60E6	6,55E6	5,46E7	1246,00	3,54E6	176302,00
3	3,91E6	2,12E6	1,93E6	7,07E6	1,62E7	3253,60	683487,00	216633,00

Conglomerado	IG_18	IG_19	IG_20	IG_21	IG_22	IG_23
1	88260,00	1755,09	67015,70	275753,00	155276,00	1,52E7
2	4,55E6	26757,00	0,00	5,36E6	4,55E6	2,39E7
3	1,54E6	14667,20	1,24E6	3,58E6	2,78E6	2,82E7

Fuente: Elaboración propia.



Fuente: Elaboración propia.

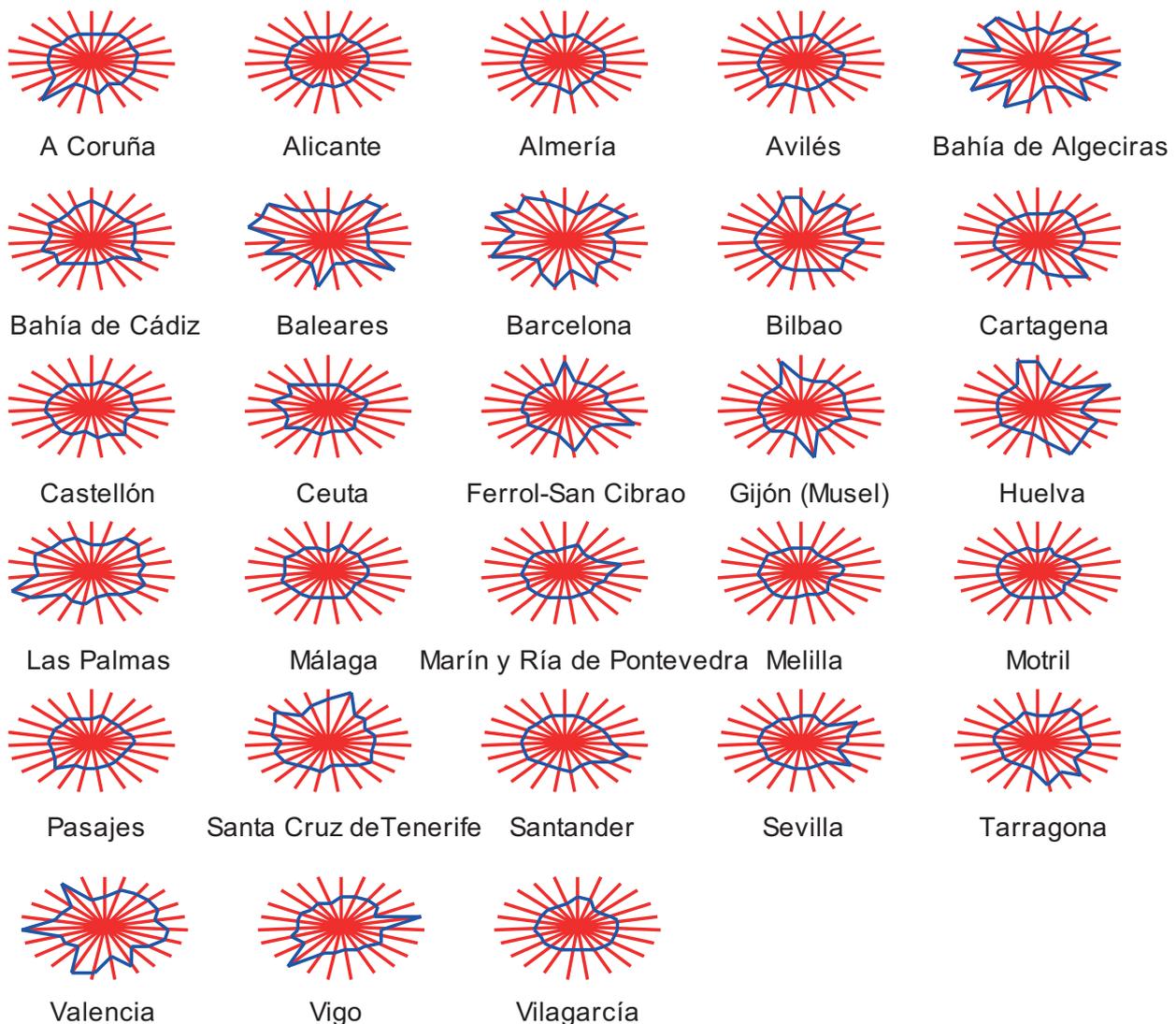
Figura 2. Dendrograma.

Con estos resultados se puede comentar en primer lugar que Algeciras aparece como elemento atípico en el conjunto de los puertos españoles, siendo el puerto mayor tráfico presente. También destaca en el tráfico de pasajeros, aunque carece de la modalidad de cruceros. El segundo grupo es el más heterogéneo y sus elementos se agrupan en último lugar. A excepción de Algeciras, corresponde a los puertos con más tráfico de España, son los puertos que tienen una mayor especialización en sus terminales y son los más relevantes en el tráfico internacional. Destacan en el tráfico de contenedores, en el avituallamiento de combustible y también en el tráfico de pasajeros, tanto en su modalidad de crucero como de tránsito.

El primer grupo es el más numeroso y el que abarca a casi el 80% de las Autoridades Portuarias (AP) españolas. Sus características son mucho más homogéneas y tienen muchas similitudes en cuanto al tamaño, tráfico nacional e internacional, especialización, etc. Este grupo se puede seguir estudiando de la misma manera. En efecto, encontramos otros tres grupos bien diferenciados. El primero, Grupo 1_1, es el más numeroso (15 AP) y homogéneo. Está compuesto por A Coruña, Bahía de Cádiz, Santander, Castellón, Sevilla, Alicante, Málaga, Almería, Avilés, Pasajes, Melilla, Motril, Vilagar-

cía, Marín y Pontevedra y Ceuta. El segundo grupo, Grupo 1_2, presenta, al igual que antes, un solo elemento (1 AP), Vigo. Por último, el tercer grupo, Grupo 1_3, resulta menos compacto y con menos elementos (6 AP): Bilbao, Cartagena, Tarragona, Ferrol-San Cibrao, Gijón y Huelva. Se puede apreciar que el Puerto de Vigo también queda como valor atípico en esta subdivisión. Las características del Puerto de Vigo hacen de él un puerto de importancia internacional en pesca y que también tiene relevancia en otras actividades del comercio marítimo formando parte de rutas internacionales de navegación.

Además del análisis de conglomerados, para profundizar más en los agrupamientos que se han generado, se han realizado otros dos análisis complementarios: análisis multivariante y de componentes principales. La Figura 3 muestra el gráfico de rayos para cada Autoridad Portuaria con los valores de todos los indicadores que se han usado. En este gráfico se pueden comparar fácilmente las Autoridades Portuarias entre sí. Por ejemplo, se puede observar claramente la semejanza entre Melilla y Motril, que son los dos primeros elementos en agruparse en el *clustering*, al igual que se aprecia la diferencia de Bahía de Algeciras con el resto de las Autoridades Portuarias.



Fuente: Elaboración propia.

Figura 3. Gráfico de rayos para las Autoridades Portuarias.

El siguiente paso ha sido obtener las componentes principales de los agrupamientos. En la Tabla 4 se muestran los porcentajes de las componentes de las variables. En este caso, se han extraído 6 componentes, puesto que 6 componentes tuvieron *eigenvalores* mayores o iguales que 1,0. En conjunto ellas explican 83,6% de la variabilidad en los datos originales.

Tabla 4. Análisis de las componentes principales.

Componente Número	Autovalor	Porcentaje de Varianza	Porcentaje Acumulado
1	9,32	40,52	40,52
2	3,07	13,38	53,90
3	2,47	10,74	64,65
4	1,86	8,10	72,75
5	1,30	5,66	78,42
6	1,20	5,22	83,64
7	0,90	3,95	87,59
8	0,76	3,31	90,91
9	0,62	2,72	93,64
10	0,42	1,86	95,50
11	0,27	1,18	96,68
12	0,26	1,13	97,81
13	0,23	1,02	98,84
14	0,13	0,58	99,42
15	0,06	0,26	99,68
16	0,04	0,18	99,86
17	0,01	0,06	99,93
18	0,00	0,02	99,96
19	0,00	0,02	99,98
20	0,00	0,01	100,00
21	0,00	0,00	100,00
22	0,00	0,00	100,00
23	0,00	0,00	100,00

Fuente: Elaboración propia.

La Tabla 5 muestra la tabla de pesos de los componentes principales, en donde los valores de las variables en la ecuación se han estandarizado restándoles su media y dividiéndolos entre sus desviaciones estándar. Por ejemplo, el primer componente principal responde a la ecuación:

Componente 1 =

$$0,15*IG_1 + 0,07*IG_2 + 0,28*IG_3 + 0,28*IG_4 + 0,17*IG_5 + 0,004*IG_6 + 0,23*IG_7 + 0,20*IG_8 + 0,20*IG_9 + 0,16*IG_{10} + 0,16*IG_{11} + 0,01*IG_{12} + 0,28*IG_{13} + 0,25*IG_{14} - 0,04*IG_{15} + 0,22*IG_{16} + 0,24*IG_{17} + 0,25*IG_{18} + 0,23*IG_{19} + 0,19*IG_{20} + 0,23*IG_{21} + 0,29*IG_{22} + 0,15*IG_{23} \quad [6]$$

Las primeras componentes son las más discriminatorias, y sólo la primera agrupa al 40% de los elementos. En el primer componente casi todos los indicadores tienen una importancia bastante similar. Las variables más significativas, en este caso todas positivas, hacen referencia a datos de calado (IG_3), longitud de muelles (IG_4), mercancía general convencional (IG_13) y vehículos en pasaje (IG_22).

Las variables más destacadas del segundo componente son las positivas de superficie de almacenamiento (IG_5), la de tráfico de productos líquidos (IG_11), graneles sólidos (IG_12) e inversiones (IG_23). Entre las negativas destaca el tráfico total de pasajeros (IG_21).

Por último, las variables más destacadas del tercer componente son las positivas de número de grúas (IG_7) y la de tráfico en contenedores (IG_14), mientras que entre las negativas destaca el número de pasajeros de cruceros (IG_20).

En la Tabla 6 se muestran los valores que toma cada Autoridad Portuaria con las 6 componentes principales más signifi-

Tabla 5. Tabla de pesos de los componentes principales.

Indicador	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
IG_1	0,15	0,27	-0,20	0,11	-0,42	-0,12
IG_2	0,07	-0,08	-0,17	0,17	-0,55	-0,34
IG_3	0,28	0,01	-0,12	0,05	0,12	0,26
IG_4	0,28	0,01	-0,12	0,05	0,13	0,26
IG_5	0,17	0,33	-0,08	-0,02	0,09	-0,30
IG_6	0,004	-0,07	-0,09	0,65	0,12	-0,001
IG_7	0,23	0,06	0,366	0,09	-0,14	0,14
IG_8	0,20	0,28	0,04	0,11	-0,10	0,02
IG_9	0,20	0,01	-0,32	0,01	-0,04	0,01
IG_10	0,16	0,23	0,16	-0,001	-0,09	0,31
IG_11	0,16	0,34	-0,02	-0,05	0,28	-0,30
IG_12	0,01	0,39	-0,02	-0,06	0,14	0,37
IG_13	0,28	-0,14	-0,06	-0,08	0,22	-0,13
IG_14	0,25	-0,04	0,36	0,09	0,04	-0,18
IG_15	-0,04	-0,08	-0,13	0,62	0,13	0,14
IG_16	0,22	-0,11	0,28	0,03	-0,19	0,14
IG_17	0,24	-0,20	-0,20	0,006	0,10	0,04
IG_18	0,25	-0,04	0,33	0,08	0,07	-0,21
IG_19	0,23	-0,26	-0,04	-0,14	-0,13	0,24
IG_20	0,19	-0,17	-0,35	-0,15	0,23	-0,12
IG_21	0,23	-0,28	-0,12	-0,16	-0,23	0,10
IG_22	0,29	-0,10	0,14	0,01	0,15	-0,23
IG_23	0,15	0,30	-0,22	-0,04	-0,15	0,05

Fuente: Elaboración propia.

Tabla 6. Tabla de componentes principales.

Autoridad Portuaria	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
A Coruña	-1,13	0,14	-0,46	1,81	0,41	0,95
Alicante	-2,51	-0,79	0,68	-0,65	0,11	-0,38
Almería	-2,31	-0,27	0,40	-0,64	0,16	0,14
Avilés	-2,33	-0,47	0,30	-0,14	0,27	-0,07
Bahía de Algeciras	8,31	-1,12	5,02	0,42	-1,53	1,36
Bahía de Cádiz	-0,45	0,30	-1,42	0,16	-0,89	-0,53
Baleares	4,38	-3,46	-3,70	-1,93	0,58	1,71
Barcelona	6,17	0,13	-1,90	-1,30	2,78	-1,55
Bilbao	2,09	2,43	-0,27	0,06	-0,55	1,04
Cartagena	-0,51	1,42	0,58	-0,45	0,73	0,89
Castellón	-1,60	0,24	0,79	-0,38	0,30	0,29
Ceuta	-2,27	-1,84	0,83	-1,11	-0,27	0,03
Ferrol-San Cibrao	-0,18	2,68	-0,01	-0,005	-0,73	0,31
Gijón (Musel)	-0,81	2,68	-0,37	-0,43	0,21	1,74
Huelva	1,34	5,22	-1,58	-0,01	-0,72	-1,42
Las Palmas	3,72	-1,46	-0,94	0,15	0,20	0,70
Málaga	-1,96	-1,19	0,28	-0,74	0,31	-0,32
Marín y Ría de Pontevedra	-2,61	-0,80	0,59	0,48	-0,11	-0,55
Melilla	-2,98	-1,17	0,80	-0,86	0,004	-0,50
Motril	-2,93	-0,78	0,78	-0,81	0,07	-0,40
Pasajes	-2,50	-0,85	0,58	0,08	0,40	-0,02
Sta. Cruz de Tenerife	2,24	-1,40	-2,51	0,37	-4,20	-1,21
Santander	-1,08	0,43	-0,08	-0,14	-0,24	0,13
Sevilla	-1,54	0,03	-0,03	-0,38	0,23	-0,92
Tarragona	0,15	1,93	0,33	0,01	0,73	1,89
Valencia	5,10	0,15	2,22	1,10	1,30	-2,94
Vigo	-1,08	-1,55	-1,45	5,88	0,93	0,37
Vilagarcía	-2,70	-0,61	0,54	-0,51	-0,53	-0,76

Fuente: Elaboración propia.

ficativas. Queda en evidencia que casi todos los puertos del Grupo 2 (Baleares, Barcelona, Las Palmas y Valencia) y el Grupo 3 (Bahía de Algeciras), son los que tienen valores más elevados del primer componente. A su vez, el componente 2 diferencia a Baleares de forma negativa y a Huelva en positivo. El componente tercero destaca Bahía de Algeciras en positivo y Baleares en negativo.

Siguiendo con estas apreciaciones, en otras componentes principales más alejadas, por ejemplo, la cuarta, destacan los indicadores IG_6 (almacenes frigoríficos) e IG_15 (pesca). Teniendo en cuenta, que existe cierta relación entre los almacenes frigoríficos y la pesca, también se observa que la Autoridad Portuaria de Vigo es la que presenta un valor más elevado de esta componente.

Finalmente, a partir del estudio de las componentes principales, no sólo se conocen los indicadores más significativos, sino también se puede considerar que los indicadores menos relevantes de los 23 elegidos son IG_2 (Zona 2) e IG_10 (productos petrolíferos), lo que permitiría prescindir de estos dos indicadores para la caracterización de los puertos.

5. CONCLUSIONES

Este trabajo responde a la necesidad de caracterizar los puertos mediante indicadores para poder realizar una clasificación entre ellos. Para ello se ha utilizado el criterio de defini-

ción genérica de los puertos, con información de los tráficos de mercancías y pasajeros, datos físicos y económicos. Se deben evitar los datos que estén interrelacionados o que no sean relevantes con el criterio que se estudia. Por ello, es recomendable utilizar aquellos indicadores consensuados con criterio experto, como se ha hecho en esta investigación.

El análisis *cluster* permite trabajar en el entorno portuario para realizar agrupamientos, comparaciones o clasificaciones entre puertos. El número de indicadores no debe ser muy alto para que la herramienta de análisis funcione en las mejores condiciones y resulta conveniente reducir este número al mínimo posible, a los más significativos. Resulta muy adecuado complementar el análisis *cluster* con otros análisis estadísticos. El análisis de componentes principales permite conocer cuáles son los indicadores más determinantes que se han usado para caracterizar los puertos. Una vez que se han identificado las principales variables en las primeras componentes principales, éstas se podrían analizar en detalle por separado, calculando estadísticos y gráficas de manera individual.

Los resultados de esta investigación aplicados al caso de los puertos españoles han respondido correctamente a la realidad del sistema portuario. En este estudio se ha tratado de una caracterización de los puertos españoles bajo un punto de vista de definición general. Con el análisis *cluster* se aprecia que los grupos no se generan según una situación geográfica definida o un frente marítimo. Sus puntos en común se refie-

ren al tamaño del puerto y a las actividades predominantes. Los puertos menos especializados son mucho más numerosos y se agrupan mucho antes que los que destacan en alguna actividad (sea pasajeros, pesca o contenedores, por ejemplo). Los puertos españoles de más relevancia internacional son los que se agrupan en último lugar, es decir, son los más heterogéneos del conjunto. La Autoridad Portuaria Bahía de Al-

geciras queda como un elemento atípico y es el último que se junta a los demás en el proceso de los agrupamientos.

Los indicadores más significativos según esta caracterización han sido: calado, muelles, mercancía general convencional y vehículos, y los menos relevantes: la superficie de Zona 2 y los productos petrolíferos.

REFERENCIAS

- (1) Caragliu, A., Del Bo, C., & Nijkamp, P. (2009). Smart cities in Europe. Vrije Universiteit, Faculty of Economics and Business Administration.
- (2) Davarzani, H., Fahimnia, M., Bell, M., & Sarkis, J. (2016). Greening ports and maritime logistics: A review. *Transportation Research Part D: Transport and Environment*, 48, 473-487.
- (3) Gutiérrez, J., Arbulo, P., & Basurto, P. (2007). Competitividad del transporte multimodal a larga distancia: Creación de un sistema de información para las decisiones estratégicas basado en indicadores. XI Congreso de Ingeniería de Organización: Madrid, 5-7 de Septiembre de 2007, 1747-1756.
- (4) Notteboom, T. E., & de Langen, P. W. (2015). Container port competition in Europe. *Handbook of ocean container transport logistics* (pp. 75-95) Springer.
- (5) Orozco, J. M. (2006). El sistema español de evaluación de impacto ambiental: Nuestra posición en el mundo. *Informes de la Construcción*, 58(504), 29-38.
- (6) Fernández-Sánchez, G., & Rodríguez-López, F. (2011). Propuesta para la integración de criterios sostenibles en los proyectos de ingeniería civil: Un caso práctico. *Informes de la Construcción*, 63(524), 65-74.
- (7) Pearson, K. (1901). LIII. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572.
- (8) Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417-441.
- (9) Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7(2), 179-188.
- (10) Jolliffe, I. T. (1986). Principal component analysis and factor analysis. *Principal component analysis* (pp. 115-128) Springer.
- (11) Tovar, B., & Rodríguez-Déniz, H. (2015). Classifying ports for efficiency benchmarking: A review and a frontier-based clustering approach. *Transport Reviews*, 35(3), 378-400.
- (12) Zemp, S., Stauffacher, M., Lang, D. J., & Scholz, R. W. (2011). Classifying railway stations for strategic transport and land use planning: Context matters! *Journal of Transport Geography*, 19(4), 670-679.
- (13) Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence*, 17(5-6), 375-381.
- (14) Asgari, N., Farahani, R. Z., & Goh, M. (2013). Network design approach for hub ports-shipping companies competition and cooperation. *Transportation Research Part A: Policy and Practice*, 48, 1-18.
- (15) Cerbán, M. M. (2008). La competitividad portuaria de los puertos de titularidad estatal. *Economistas*, (116), 277-282.
- (16) Laxe, F. G., Seoane, M. J. F., & Montes, C. P. (2015). Política y evaluación de la selección portuaria: El caso español. *Revista de Evaluación de Programas y Políticas Públicas*, 1(4), 82-102.
- (17) González, N., Soler, F., & Camarero, A. (2013). Modelo de eficiencia de las terminales de contenedores del sistema portuario español. *Rect@: Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, (14), 49-67.
- (18) Borén, S., Nurhadi, L., Ny, H., Robèrt, K., Broman, G., & Trygg, L. (2017). A strategic approach to sustainable transport system development—part 2: The case of a vision for electric vehicle systems in southeast Sweden. *Journal of Cleaner Production*, 140, 62-71.
- (19) Serrano, O. (2015). Operativa portuaria y sostenibilidad. *CONAMA LOCAL*, 7.
- (20) Sánchez, R., Jaimurzina, A., Wilmsmeier, G., Pérez, G., Doerr, O., & Pinto, F. (2015). Transporte marítimo y puertos: Desafíos y oportunidades en busca de un desarrollo sostenible en América Latina y el Caribe.
- (21) Duru, O. (2017). The origin and consistency of the Ton-Mile metric in the shipping economics. *Logistics*, 1(1), 3.
- (22) Rodríguez, A., & Laio, A. (2014). Machine learning. clustering by fast search and find of density peaks. *Science* (New York, N.Y.), 344(6191), 1492-1496. doi: 10.1126/science.1242072
- (23) Puertos del Estado. (2015). In Organismo Público Puertos del Estado (Ed.), *Memorias anuales de puertos del estado*
- (24) Camarero, A., & González, N. (2005). Cadenas integradas de transporte. Fundación Agustín de Betancourt. Ministerio de Fomento.
- (25) Camarero, A., & González, N. (2007). Logística y transporte de contenedores. Fundación Agustín de Betancourt. Ministerio de Fomento.

* * *