

ANÁLISIS COMPARATIVO DE LAS PROVINCIAS DE CASTILLA Y LEÓN SEGÚN SUS EXPORTACIONES E IMPORTACIONES EN EL PERIODO 1990-96

M^a Carmen HUERGA CASTRO

M^a Eva VALLEJO PASCUAL

Julio ABAD GONZÁLEZ

Área de Estadística e I.O. Universidad de León

1. INTRODUCCIÓN

El objetivo de este estudio es analizar las provincias castellano-leonesas en el período 90-96 en función del tipo de productos exportados e importados. En este sentido, se tratará de encontrar semejanzas y diferencias entre las provincias de nuestra Comunidad Autónoma, que puedan ser consecuencia de la distinta composición de su tejido industrial.

La aplicación de las técnicas *STATIS* y *cluster* permitirán alcanzar dicho objetivo, poniendo al mismo tiempo de manifiesto sus posibilidades a la hora de analizar los diversos aspectos de la realidad económica.

En primer lugar se realizará una breve descripción teórica de ambas técnicas, siendo aplicadas posteriormente a los datos objeto de análisis. Las conclusiones obtenidas y la principal bibliografía empleada constituyen el final de la presente comunicación.

2. METODOLOGÍA

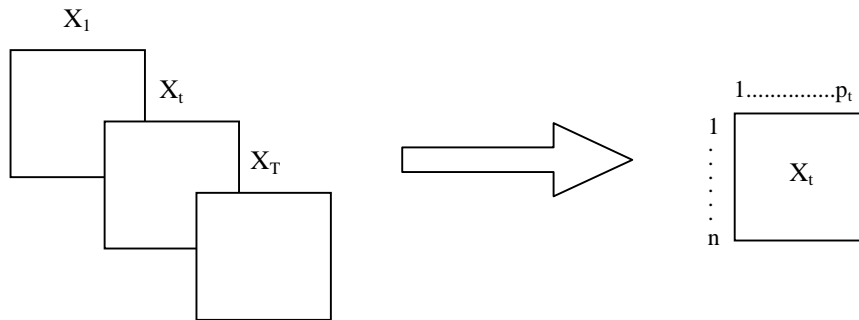
2.1. Método STATIS

La técnica *Statis* permite el tratamiento de varias tablas de datos simultáneamente. Esta metodología comenzó a desarrollarse en los años 70 por el equipo investigador del profesor ESCOUFIER, recogida en la tesis doctoral de L'HERMIER DES PLANTES (1976), y basada en desarrollos matemáticos anteriores. Posteriormente ha sido desarrollada por varios autores entre los que cabe citar a LAVIT (1988) y PÉREZ-HUGALDE (1987) y ha sido incluida en paquetes estadísticos tales como SPAD o ADE-4.

Partimos de T tablas de datos $X_1, X_2, \dots, X_t, \dots, X_T$, cada una de las cuales representa n individuos (filas) sobre los que se han medido p_t variables cuantitativas

(columnas). La tabla total es una matriz de n filas y $\sum_{t=1}^T p_t$ columnas. A menudo las tablas representan T diferentes momentos del tiempo en los cuales se han observado un conjunto n de individuos respecto a p_t características (las mismas o diferentes en cada periodo).

Figura 1. Notación de las tablas de datos



La técnica multivariante empleada para el tratamiento de este tipo de tablas de datos es el Análisis de Componentes Principales (ACP). Hay que precisar que el ACP de la tabla total (de las $\sum_{t=1}^T p_t$ columnas yuxtapuestas) no es válido en la medida en que no considera que las variables pertenecen a grupos o tablas diferentes, y que a menudo (cuando se trata de datos temporales) están correlacionadas de un periodo a otro. Por otra parte, un ACP de cada grupo o tabla no permite comparar adecuadamente dichos grupos en la medida en que cada uno de ellos tiene ejes factoriales diferentes.

En cambio, el método STATIS (Structuration des Tableaux À Trois Indices de la Statistique) permite:

- * La comparación global de las nubes de individuos que constituye cada tabla, en la etapa denominada *estudio de la interestructura*. Se trata de determinar en qué medida las tablas tienen estructuras semejantes o no, a partir de la imagen euclídea plana obtenida, en la que cada punto representa una tabla de datos. Para el caso en el que éstas representen el tiempo, la semejanza indicará que los individuos que las constituyen mantienen su comportamiento en el tiempo respecto a las variables objeto de estudio.
- * La obtención de una nube de individuos "media" o *matriz compromiso*, resumen de las diferentes tablas, que representa la estructura común a dichas tablas.

- * Análisis de la matriz "compromiso", en la etapa denominada *estudio de la infraestructura*, que permite representar los individuos medios del conjunto de las tablas (individuos-compromiso), así como las variables de cada tabla con el fin de caracterizar los ejes obtenidos e interpretar las posiciones de esos individuos.
- * *Representación simultánea o representación de las trayectorias* de los individuos en la imagen euclídea del compromiso. El estudio de éstas permitirá conocer la evolución de cada uno de los individuos, así como los responsables de las diferencias entre tablas.

Expondremos brevemente los aspectos teóricos más importantes del método. Desarrollos matemáticos más detallados están recogidos en GLAÇON (1981) y LAVIT (1988). Dado que las etapas señaladas se basan en un ACP, comenzaremos por una breve descripción de esta técnica.

Para una tabla X_t (de n filas y p_t columnas), consideramos la matriz diagonal D de pesos de los individuos, común a todas las tablas, y la matriz diagonal M_t de pesos de las variables (que coincide con la matriz identidad para el caso más general, en el que todas las variables tiene el peso unitario). Cada una de las columnas de la matriz X_t está estandarizada para que tenga media cero.

La matriz $V_t = X_t' D X_t M_t$, de orden $(p_t \times p_t)$, es la matriz de covarianzas (o de correlaciones si las variables han sufrido otro cambio de escala para tener varianza unitaria).

La matriz $W_t D = X_t M_t X_t' D$, de orden $(n \times n)$, es la matriz de productos escalares entre individuos. Ambas matrices tienen los mismos autovalores.

Si los autovectores de V_t son las columnas de una matriz U , entonces las columnas de $X_t U$ son los vectores propios de $W_t D$, normalizados por los correspondientes autovalores.

Llamando Λ a la matriz diagonal de los autovalores en orden decreciente, se cumple que:

$$U' U = I_{(p_t \times p_t)} \quad U' X_t' D X_t U = \Lambda_{(n \times n)}$$

Las columnas de U son los ejes principales y las columnas de $X_t U$ son las coordenadas de los n individuos en los correspondientes ejes principales (componentes principales). La representación gráfica así obtenida del primer y segundo ejes principales permite obtener la nube de individuos en un espacio de menor dimensiones, que es el que mejor se ajusta a la forma original de la nube de puntos.

Las columnas de $G = X_t' D X_t U \Lambda^{-1/2}$ representan las coordenadas de las variables en la base ortonormal de las columnas de $X_t U \Lambda^{-1/2}$. Esas coordenadas representan la correlación con los ejes (si las variables tienen varianza unitaria) y permiten su interpretación.

Pasamos a desarrollar el método STATIS, propiamente dicho empezando por *el estudio de la interestructura*. Como ya hemos señalado, sólo vamos a considerar el caso particular de que las variables de cada matriz X_t están medidas sobre el mismo conjunto de n individuos. Las variables no tiene porque ser las mismas en todas las tablas.

La matriz $W_t = X_t M_t X_t'$ es simétrica y semi-definida positiva. Así, el operador $W_t D$ es una matriz de productos escalares entre individuos y el objeto representativo de cada tabla t .

Para comparar dichos objetos a través de su imagen euclídea, es necesario definir una distancia entre ellos. Dicha distancia se deduce del producto escalar de HILBERT-SCHMIDT entre aplicaciones lineales. Así, según ROBERT y ESCOUFIER (1976), el producto escalar:

$$\langle W_t D, W_{t'} D \rangle = \text{tr}(W_t D W_{t'} D) = \text{cov}(W_t D W_{t'} D)$$

y en el caso de operadores normados (cuando las normas son muy diferentes entre si, se equilibra la influencia de cada tabla dividiendo entre la norma del operador):

$$\left\langle \frac{W_t D}{\|W_t D\|}, \frac{W_{t'} D}{\|W_{t'} D\|} \right\rangle = \frac{\text{tr}(W_t D W_{t'} D)}{\sqrt{\text{tr}(W_t D)^2} \sqrt{\text{tr}(W_{t'} D)^2}} = RV(W_t D, W_{t'} D)$$

La matriz C –de orden $(T \times T)$ – cuyos elementos son los productos escalares $\text{cov}(W_t D W_{t'} D)$, se puede interpretar como una matriz de covarianzas. La representación de las T tablas se obtiene mediante el ACP de C , siendo las coordenadas de las tablas los componentes de los autovectores de C multiplicados por la raíz cuadrada del valor propio correspondiente.

La matriz C se puede interpretar como una matriz de correlaciones entre los operadores cuando tomamos como elementos de C los $RV(W_t D, W_{t'} D)$, y los puntos representativos de cada tabla dan lugar a un círculo de radio unitario. Por ello, el producto RV puede interpretarse como el coseno del ángulo entre operadores.

Si las variables tienen media cero la $tr(W_t D W_t D) = \sum_j^{p_t} \sum_{j'}^{p_{t'}} cov^2(X_j^t, X_{j'}^{t'})$, y si

las variables tienen varianzas unitarias entonces la traza vale $\sum_j^{p_t} \sum_{j'}^{p_{t'}} corr^2(X_j^t, X_{j'}^{t'})$. Por

tanto, cuanto más cerca están $W_t D$ y $W_{t'} D$ en el gráfico (menor ángulo), más próximas estarán X_t y $X_{t'}$ en términos relativos.

Dado que C es positiva (es el resultado de productos escalares semi-definidos positivos) todos los coeficientes de su primer autovector pueden ser elegidos positivos (teorema de FROBENIUS). Por ello, tomando ese vector propio estandarizado

$(\alpha_1, \dots, \alpha_T)$, $\sum_{t=1}^T \alpha_t^2 = 1$, el objeto compromiso $WD = \sum_{t=1}^T \alpha_t W_t D$, que es una

combinación lineal positiva de matrices semi-definidas positivas, es también semi-definida positiva. La matriz WD se denomina *matriz compromiso*, y representa una matriz ficticia de productos escalares entre individuos que resumen lo mejor posible las $W_t D$. Así, WD es colineal al primer eje del gráfico de la interestructura y si cualquier $W_t D$ es ortogonal a ese eje, dicho $W_t D$ tendrá un coeficiente bajo en la definición del compromiso.

Se cumple que $\sum_{t=1}^T \langle WD, W_t D \rangle^2 = \lambda_1$ y que $\|WD\|^2 = 1$.

El ACP de esa matriz compromiso (*estudio de la intraestructura*) es equivalente a realizar un ACP de las tablas yuxtapuestas $X = (\sqrt{\alpha_1} X_1, \dots, \sqrt{\alpha_T} X_T)$. Dicho análisis permite, por consiguiente, obtener la representación simultánea de las variables de las T tablas en una base D -ortonormal de las columnas de $XU\Lambda^{-1/2}$ por $X'DXU\Lambda^{-1/2}$ (las columnas de $\sqrt{\alpha_t} X_t' D X U \Lambda^{-1/2}$ son las coordenadas de las $\sum_{t=1}^T p_t$ variables de la tabla

T). Además las columnas de XU son las coordenadas de los individuos-compromiso. En la mayoría de los casos se pueden explicar los ejes mediante las variables con las que están correlacionados y, por tanto, interpretar la posición de dichos individuos-compromiso en los ejes.

La *representación simultánea* de los individuos no es posible con el ACP anterior. Por ello, es necesario considerar los individuos de cada tabla como suplementarios y sus coordenadas vendrán dadas por las columnas de $W_t D X U \Lambda^{-1}$. Así, los puntos que definen la trayectoria de un individuo tienen como baricentro la posición compromiso de este individuo. El punto t de la trayectoria del individuo i , sitúa a éste comparándolo con el individuo medio del momento t . El segmento que une t y t' muestra

la variación del individuo de la tabla t a la tabla t' , con relación a la variación media. Se distinguen dos casos:

- * una trayectoria poco extensa y entorno a su posición compromiso corresponde a un individuo cuya evolución sigue la evolución media, es decir, para cada variable observada, la desviación entre el valor de la variable y la media es regular de unos años a otros.
- * una trayectoria de gran amplitud refleja un cambio de estructura de los individuos a lo largo de los años, diferente a la evolución media.

Si además es posible interpretar los ejes del compromiso por medio de las variables, la interpretación de esa trayectoria es más sencilla.

2.2. Análisis Cluster

La denominación *análisis cluster* se utiliza para definir una serie de técnicas estadísticas, fundamentalmente algoritmos, que permiten la obtención de grupos de elementos similares. Consiste en clasificar una serie de individuos, de los que se dispone de una serie de observaciones, en grupos lo más homogéneos posible en base a las variables observadas. El análisis cluster se denomina también análisis de conglomerados, análisis tipológico, o clasificación automática, entre otros sinónimos.

El punto de partida del análisis cluster es, en general, una matriz de datos formada por los valores de una serie de variables (que pueden ser cuantitativas y/o cualitativas) medidas sobre cada uno de los individuos que constituyen el objeto del estudio. Se puede utilizar para agrupar tanto individuos como variables (el proceso es idéntico en ambos casos). Las etapas de un análisis cluster son las siguientes:

- * *Selección de las variables.* Consiste en elegir las características o variables relevantes para identificar los grupos o clusters de acuerdo con los objetivos que se pretendan en el estudio. El número de variables elegidas no debe ser excesivo y las unidades en que se miden han de ser homogéneas.
- * *Elección de la medida de asociación.* La determinación de los grupos lleva implícita una medida que permita analizar la proximidad entre los individuos o entre las variables. Como principales medidas de asociación entre variables señalamos el coseno del ángulo de los vectores de dos variables, y el coeficiente de correlación entre dos variables genéricas medidas sobre los individuos objeto de estudio. Para analizar la proximidad de los individuos, las medidas de asociación más utilizadas son la distancia euclídea, la distancia de Minkowski y la de Mahalanobis.

- * *Elección de la técnica o criterio de agrupación.* Depende de la naturaleza de los datos y de los objetivos perseguidos. Se distinguen los métodos jerárquicos y no jerárquicos.
- * *Validación e interpretación de resultados.* Existen diversos métodos para comprobar en qué medida la estructura final obtenida es válida. Se suele utilizar el coeficiente de correlación cofenético cuando se han utilizado métodos jerárquicos. En el caso de los no jerárquicos algunos autores proponen utilizar el análisis multivariante de la varianza o bien desarrollar análisis de la varianza sobre cada variable en cada cluster. En cuanto a la interpretación de resultados hay que señalar que ninguna de las técnicas proporciona una solución óptima para todos los problemas, ya que es posible llegar a distintos resultados según la técnica elegida. Por esta razón, el criterio del investigador, el conocimiento del problema y la experiencia deben sugerir el método adecuado y la solución correcta.

Como ya hemos señalado, los métodos cluster se pueden clasificar en dos grandes grupos: los **jerárquicos** y los **no jerárquicos**. Estos últimos tratan de realizar una única partición de los individuos en k grupos, a partir de la matriz de datos original. Por el contrario, los métodos jerárquicos van construyendo clusters sucesivamente de forma que se minimice alguna distancia o se maximice alguna medida de similitud. Asimismo, permiten la construcción de un árbol de clasificación, denominado dendograma, en el que se observa qué grupos se van uniendo, en qué nivel y qué valor toma la medida de asociación.

Cabe distinguir dos tipos de métodos jerárquicos: los **aglomerativos** o **ascendentes** y los **disociativos** o **descendentes**. Los últimos comienzan con un conglomerado que engloba a todos los casos y, a través de sucesivas divisiones, se forman grupos cada vez más pequeños y homogéneos. En cambio, los métodos aglomerativos comienzan con tantos grupos como individuos haya y se van formando grupos, de forma ascendente, hasta que al final todos los casos están englobados en un mismo conglomerado.

Entre los métodos aglomerativos señalamos los de la distancia mínima entre grupos o similitud máxima y los de la distancia máxima o similitud mínima. El método de promedio inter-grupos define la distancia entre dos clusters como el promedio de las distancias entre todos los posibles individuos de los clusters formados anteriormente. El método de la media ponderada combina los grupos con menor distancia promedio. Existen otros métodos, como los basados en el centroide, que se caracterizan porque la semejanza entre dos clusters viene dada por la semejanza entre sus centros de gravedad, es decir, los vectores de medias de las variables medidas sobre los individuos del cluster. Destacan entre ellos, el método del centroide ponderado y el de Ward. Este último obtiene la suma de las distancias euclídeas al cuadrado entre cada individuo y la media de

las variables de su grupo. Los clusters que se forman son los que resultan con el menor incremento en la suma total de las distancias al cuadrado intra-cluster

Un análisis cluster puede acompañarse de técnicas factoriales. Algunos algoritmos de análisis cluster comienzan con un ACP para reducir el número de variables originales a un número menor de componentes principales, aunque obviamente los resultados obtenidos pueden ser distintos. Se pueden tomar las coordenadas de los individuos en los primeros ejes factoriales como las nuevas variables, consiguiendo así un resumen de la tabla de datos inicial. Por otro lado, cuando las dos primeras componentes principales explican un alto porcentaje de inercia, la representación gráfica de los individuos en el plano de ambas componentes, permite una primera aproximación para buscar clusters. De esta forma un análisis factorial se presenta como un filtro que preserva la información inicial más útil.

3. APLICACIÓN

Los datos empleados son los recogidos en el Anuario Estadístico de Castilla y León correspondiente al año 1997 (tablas 9.2.3 y 9.2.4).

Para cada uno de los años 1990, 1992, 1994 y 1996, se han tomado las exportaciones y las importaciones –desagregadas en nueve tipos de productos y en millones de pesetas– de cada una de las provincias de Castilla y León. Se analizarán, por separado, exportaciones e importaciones empleando el método STATIS sobre las tablas centradas en columnas. Con las puntuaciones factoriales de los individuos-compromiso, resultado del análisis de la infraestructura, tanto de importaciones como de exportaciones, realizaremos un análisis cluster, y clasificaremos las provincias castellano-leonesas. Los programas estadísticos empleados son SPAD y ADE-4.

Tanto la tabla de importaciones como la de exportaciones está compuesta por cuatro subtablas (una por cada año considerado) de nueve individuos-filas (las provincias) y nueve columnas-variables (los nueve tipos de productos considerados). La notación empleada se recoge a continuación:

Tabla 1. Notación de variables(*) e individuos

VARIABLES		INDIVIDUOS	
PA	Productos agropecuarios	Av	Ávila
PQ	Productos químicos	Bu	Burgos
PM	Papes y sus manufacturas	Le	León
MC	Metales comunes y sus manufacturas	Pa	Palencia
MN	Maquinaria no eléctrica	Sa	Salamanca
ME	Maquinaria eléctrica	Se	Segovia
MT	Material de transporte	So	Soria
IO	Instrumentos de óptica y fotografía	Va	Valladolid
OT	Otros	Za	Zamora

(*) La clasificación de los productos se realiza según nomenclatura comunitaria.

En aquellos casos en los que tanto una variable como un individuo se refieran a un año concreto, a la abreviatura del nombre le precederá el 0, el 2, el 4, el 6, según sea el año 90, 92, 94 ó 96.

3.1. Aplicación del método STATIS a la tabla de importaciones

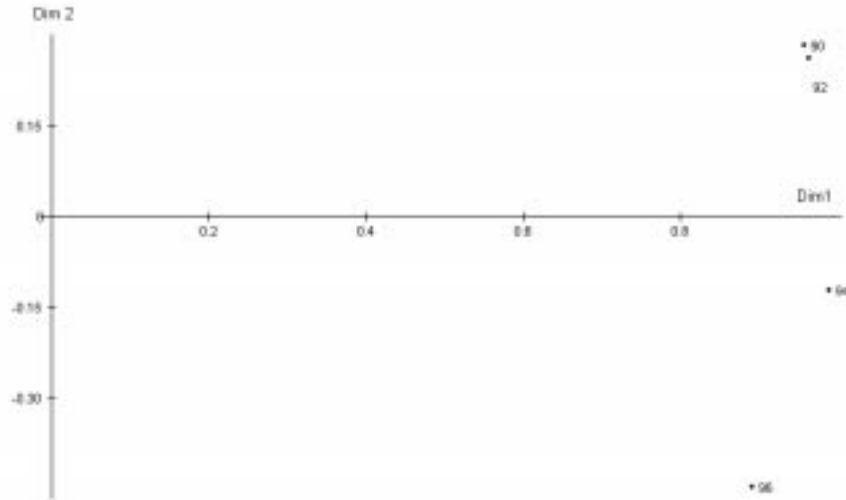
Siguiendo las etapas expuestas en el epígrafe teórico, empezaremos con *el estudio de la interestructura*. La matriz de coeficientes *RV* entre tablas, que se recoge a continuación, se puede interpretar como una matriz de correlaciones.

Tabla 2. Coeficientes *RV* (importaciones)

	90	92	94	96
90	1.000	-	-	-
92	0.999	1.000	-	-
94	0.914	0.923	1.000	-
96	0.730	0.744	0.938	1.000

Los coeficientes próximos a 1 indican estructuras semejantes (entre el año 90 y 92, por ejemplo). Un valor que se aleja de 1, tal es el caso del año 96 con el resto, indicará que el comportamiento de los individuos (las provincias) de dicha tabla se ha modificado respecto a los anteriores (estructuras diferentes de las tablas). La extracción de los autovalores y autovectores de dicha matriz, tal y como se ha expuesto teóricamente, permite obtener la siguiente representación gráfica:

Gráfico 1. Imagen euclídea aproximada entre objetos (importaciones)



Cada tabla anual está representada por un punto del plano factorial, formado por el primer y segundo eje, que representa el 99,86% de la inercia total. Para su correcta interpretación, es necesario analizar la norma de cada vector y el ángulo que forma con el resto de vectores. Dado que las normas son similares y están alejadas del origen, las cuatro tablas están bien representadas y la ponderación en la construcción del compromiso será similar.

El ángulo entre los vectores representativos de los años 90 y 92 es pequeño, mientras que entre estos dos últimos y los años 94 y 96 es más grande. Si se realiza un ACP de cada tabla individual se puede comprobar que las cuatro nubes de puntos tienen una dimensión y forma similares (tienen el mismo número de valores propios no nulos y decrecen de forma similar). Es por esto, que la variación en el ángulo entre tablas que presenta la gráfica proviene de un cambio de posición de los individuos en el interior de las tablas de los años 94 y 96. En este caso, un cambio de comportamiento de las provincias respecto al tipo de productos importados en esos dos años.

El estudio de la *intraestructura* y la *representación simultánea* permitirá poner de manifiesto los individuos responsables de tales diferencias y respecto a qué variables. La construcción de la *nube de puntos compromiso* es el resultado de la siguiente combinación lineal:

$$0,264 \frac{W_1 D}{\|W_1 D\|} + 0,266 \frac{W_2 D}{\|W_2 D\|} + 0,273 \frac{W_3 D}{\|W_3 D\|} + 0,246 \frac{W_4 D}{\|W_4 D\|}$$

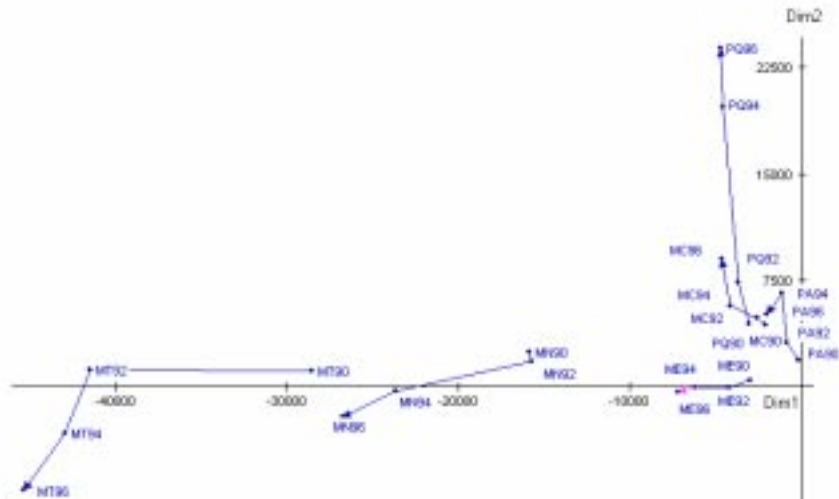
La descomposición en valores y vectores propios de la *matriz compromiso* WD permitirá obtener una representación gráfica plana, y la correlación entre cada variable y los ejes del compromiso (tabla 3) ayudará a dotarlos de significado económico.

Tabla 3. Correlaciones con los primeros ejes del compromiso (importaciones)

Variable	dim 1	dim2	dim 3	Variable	dim1	dim2	dim3
IO90	-0.964	0.150	0.212	MT94	-0.997	-0.077	-0.003
IO92	-0.977	0.063	0.184	MT96	-0.957	-0.154	-0.244
IO94	-0.996	-0.047	0.063	OT90	-0.799	0.545	0.198
IO96	-0.946	-0.041	-0.304	OT92	-0.803	0.549	0.150
MC90	-0.422	0.868	-0.043	OT94	-0.842	0.526	0.093
MC92	-0.456	0.857	-0.022	OT96	-0.774	0.628	0.014
MC94	-0.572	0.768	-0.110	PA90	-0.084	0.825	-0.035
MC96	-0.434	0.841	-0.143	PA92	-0.22	0.810	-0.103
ME90	-0.97	0.154	0.180	PA94	-0.149	0.886	-0.066
ME92	-0.979	-0.023	0.119	PA96	-0.311	0.775	-0.073
ME94	-0.999	-0.008	0.013	PM90	-0.229	0.954	-0.038
ME96	-0.988	-0.059	-0.132	PM92	-0.499	0.854	-0.036
MN90	-0.963	0.147	0.226	PM94	-0.629	0.729	0.089
MN92	-0.98	0.109	0.169	PM96	-0.352	0.871	-0.149
MN94	-0.999	-0.012	-0.045	PQ90	-0.484	0.702	0.096
MN96	-0.952	-0.076	-0.296	PQ92	-0.432	0.857	0.057
MT90	-0.967	0.037	0.253	PQ94	-0.224	0.963	-0.098
MT92	-0.968	0.027	0.248	PQ96	-0.193	0.972	-0.135

Según dicha tabla, el primer eje, que representa el 82,30% de la varianza total, está correlado negativamente con todas las variables, sobre todo con la maquinaria eléctrica (ME), maquinaria no eléctrica (MN) y material de transporte (MT). El segundo eje explica el 13,08% de la inercia y está correlado positivamente con los metales comunes (MC), el papel y las manufacturas (PM), los productos agropecuarios (PA) y los productos químicos (PQ). El gráfico 2 recoge el plano principal del compromiso, que explica el 95,38% de la varianza. Para una mejor comprensión de la misma se han eliminado ciertas variables muy cercanas al origen y que, por tanto, no permiten caracterizar los ejes.

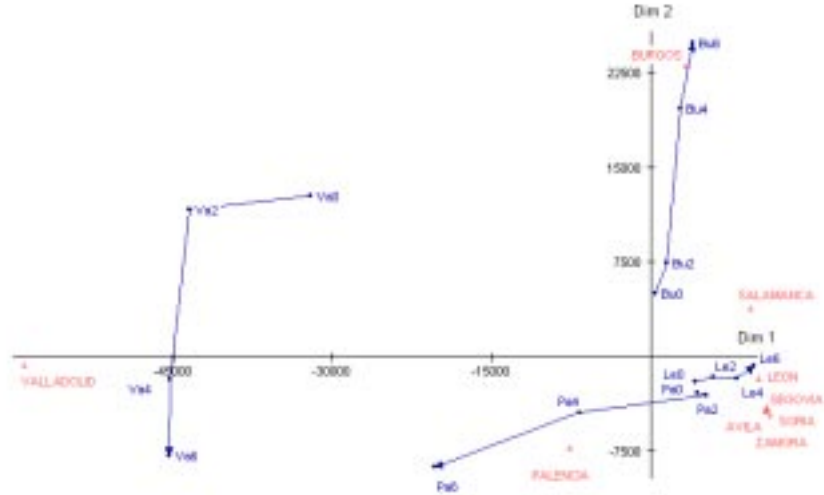
Gráfico 2. Ejes 1 y 2 del compromiso (importaciones)



La variable otros productos (OT) presenta una alta correlación con los dos ejes, si bien no se presenta en el gráfico puesto que se sitúa muy cerca del origen.

Como se ve en el gráfico 3, las posiciones compromiso con una interpretación más clara son Valladolid y Burgos. La primera asociada al primer eje (material de transporte y maquinaria no eléctrica) y Burgos al segundo eje (relacionado con los productos químicos, productos agropecuarios y metales comunes). El resto de los individuos-compromiso, no tienen una interpretación clara en cuanto que están próximos al origen. De todos modos, por su posición es evidente que tienen un comportamiento diferente a las ya mencionadas Burgos y Valladolid. Se podrían estudiar otras dimensiones, pero la varianza que explican es muy pequeña y no aportan apenas información.

Gráfico 3. Individuos-compromiso y trayectorias (importaciones)



En el gráfico 3 sólo se recogen las trayectorias de algunos individuos. Todas estas trayectorias que se modifican a lo largo del tiempo, no giran entorno a las posiciones compromiso, lo que indica que el comportamiento de esas provincias ha variado en el periodo de tiempo considerado y respecto a las variables consideradas. Estos individuos serían responsables de que las tablas del 94 y 96 presenten una estructura diferente al resto:

- * Valladolid ha ido modificando su posición respecto al segundo eje (disminución de las importaciones de productos químicos, agropecuarios y metales comunes) y un mantenimiento respecto al primer eje (excepto para el primer año).
- * Burgos va modificando su trayectoria desplazándose en torno al segundo eje a medida que transcurre el tiempo (supondría un aumento de las importaciones de productos químicos y agropecuarios).
- * Palencia ha sufrido un desplazamiento respecto al primer eje, pasando de la parte positiva del eje a la parte negativa, por lo que la posición compromiso tiene una difícil interpretación.
- * León se ha ido desplazando, aunque levemente, hacia la derecha, respecto el primer eje, lo que supondría una disminución de las importaciones de material de transporte y maquinaria eléctrica y no eléctrica. Si bien, dado que su trayectoria está próxima al origen, no podemos extraer grandes conclusiones.

3.2. Aplicación del método STATIS a la tabla de exportaciones

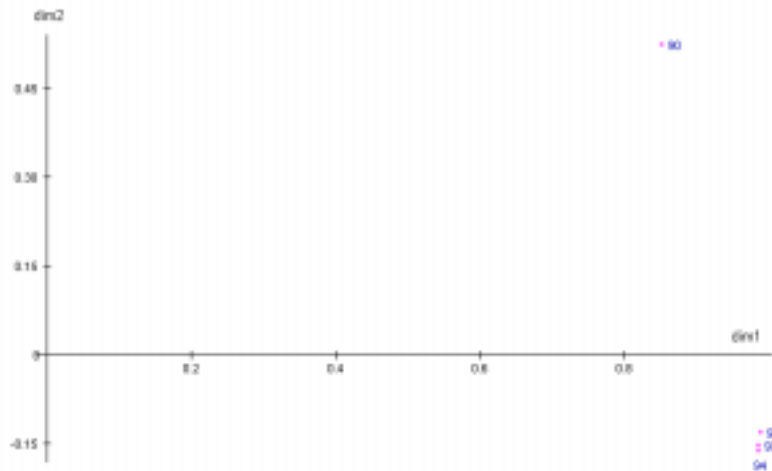
Vamos a seguir las mismas etapas que en el caso de las importaciones. El estudio de la interestructura se realizará observando la matriz de coeficientes RV (tabla 4) y el gráfico que resulta de descomponer ésta en sus vectores y valores propios (gráfico 4).

Tabla 4. Coeficientes RV (exportaciones)

	90	92	94	96
90	1.000	-	-	-
92	0.761	1.000	-	-
94	0.755	0.999	1.000	-
96	0.775	0.998	0.999	1.000

De la interpretación de esta matriz se deduce que las tablas, excepto la que representa al año 90, están próximas entre sí, es decir, conservan su estructura en el tiempo. Si se observa la representación gráfica de los dos ejes principales (99,93% de varianza explicada), se llega a la misma conclusión en cuanto que el ángulo entre los vectores que representan a los años 92, 94, y 96 es muy pequeño y los vectores son casi colineales. Por tanto, en términos económicos podemos decir que las provincias han mantenido su comportamiento respecto al tipo de productos exportados en el periodo 92-96. Puesto que todos los puntos están lejos del origen, todas las tablas están bien representadas y el compromiso es un buen resumen de todas ellas.

Gráfico 4. Imagen euclídea aproximada entre objetos (exportaciones)



La construcción de la *matriz compromiso*, que se obtiene como combinación lineal de las matrices de productos escalares de los diferentes tablas anuales, es el resultado de la siguiente expresión:

$$0,233 \frac{W_1 D}{\|W_1 D\|} + 0,270 \frac{W_2 D}{\|W_2 D\|} + 0,270 \frac{W_3 D}{\|W_3 D\|} + 0,271 \frac{W_4 D}{\|W_4 D\|}$$

Dado que el comportamiento de la matriz de productos escalares del primer año tiene una estructura diferente, su influencia (y por tanto su coeficiente) en la construcción de la matriz compromiso es algo menor. No obstante, su coeficiente sigue estando bastante próximo al resto.

La descomposición en valores singulares de la *matriz compromiso* (WD) permite obtener una imagen euclídea plana de las variables y de los individuos-compromiso. Así, el plano principal (ejes primero y segundo), representan el 95,87% de la información contenida en la tabla. A continuación se muestra la matriz de correlaciones (tabla 5) de las variables para una mejor interpretación de dichos ejes.

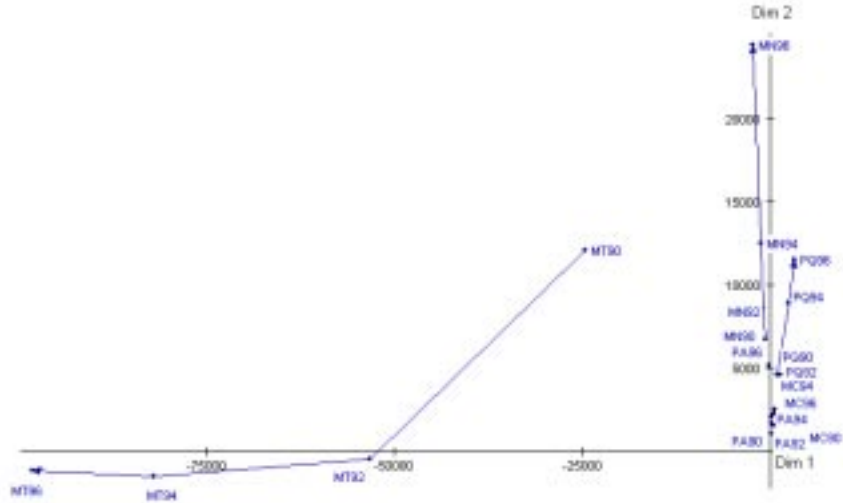
Tabla 5. Correlaciones con los primeros ejes del compromiso (exportaciones)

VARIA	DIM 1	DIM2	DIM 3
IO90	-0.329	-0.014	0.287
IO92	0.159	-0.091	-0.193
IO94	0.144	0.004	-0.196
IO96	0.165	-0.101	-0.190
MC90	0.145	-0.515	0.821
MC92	0.105	-0.666	0.721
MC94	0.113	-0.479	0.776
MC96	0.106	-0.383	0.811
ME90	0.143	-0.496	0.121
ME92	0.308	0.234	0.039
ME94	0.169	-0.324	0.865
ME96	0.196	-0.34	0.837
MN90	-0.081	-0.931	-0.333
MN92	-0.101	-0.923	-0.366
MN94	-0.089	-0.923	-0.374
MN96	-0.086	-0.917	-0.388
MT90	-0.882	-0.436	-0.174
MT92	-0.999	0.009	0.020
MT94	-1	0.019	0.011
MT96	-1	0.012	0.012
OT90	0.158	-0.33	0.915
OT92	0.15	-0.372	0.908
OT94	0.225	-0.395	0.720
OT96	0.206	-0.45	0.793
PA90	0.134	-0.696	0.595
PA92	0.089	-0.778	0.498
PA94	0.055	-0.675	0.467
PA96	-0.022	-0.892	0.292
PM90	0.132	-0.317	0.873
PM92	0.209	-0.285	0.844
PM94	0.112	-0.272	0.878
PM96	-0.195	-0.384	0.806
PQ90	0.135	-0.676	0.709
PQ92	0.166	-0.531	0.811
PQ94	0.157	-0.571	0.803
PQ96	0.152	-0.553	0.818

El primer eje presenta una alta correlación con las variables representativas del material de transporte (MT), en la parte negativa. El eje segundo está correlado, en la parte positiva, con las variables maquinaria no eléctrica (MN), productos agropecuarios

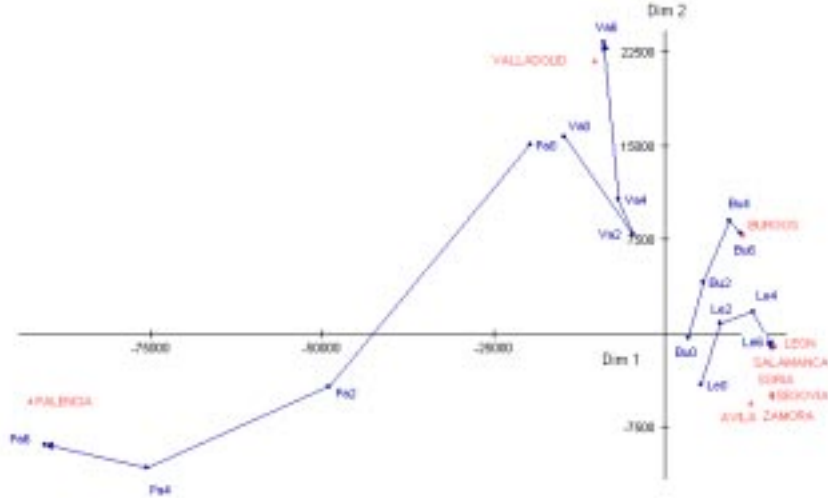
(PA) y productos químicos (PQ), aunque éstos en menor medida. Una interpretación análoga se deduce de la observación del gráfico 5.

Gráfico 5. Ejes 1 y 2 del compromiso (exportaciones)



El posicionamiento de las provincias (individuos-compromiso) en dichos ejes permitirá caracterizar los individuos en ese periodo (gráfico 6). Así, Valladolid (por su posición con respecto al segundo eje) estará caracterizada por la exportación de maquinaria no eléctrica y de productos químicos, al igual que Burgos. Por su parte, Palencia se caracterizará por la exportación de material de transporte. El resto de provincias, por su proximidad al origen, no tiene una interpretación clara aunque su comportamiento difiere del de las provincias comentadas.

Gráfico 6. Individuos-compromiso y trayectorias (exportaciones)



El estudio de las trayectorias permitirá puntualizar estas conclusiones y observar el comportamiento en el periodo objeto de estudio de algunas de estas provincias. Sólo se han trazado las trayectorias con una interpretación más clara.

Todas las trayectorias son longitudinales, lo que indica que las provincias han ido variando su posición, respecto al tipo de productos exportados, en el periodo considerado: Palencia respecto a la exportación de material de transporte (se ha desplazado hacia la izquierda afianzando más su posición de exportadora de este tipo de productos); Valladolid y Burgos, respecto a la exportación de productos químicos y de maquinaria no eléctrica (desplazamiento a lo largo del segundo eje en la parte positiva).

La trayectoria de León no tiene una interpretación clara, aunque dado que su trayectoria no es muy larga, podemos decir que su comportamiento no se ha modificado mucho en el periodo considerado.

3.3. Aplicación del análisis cluster

Una vez finalizado el análisis de importaciones y exportaciones mediante la técnica STATIS, tomamos como datos de partida para el análisis cluster los resultados obtenidos con dicho método. En concreto, vamos a tomar como variables de entrada del cluster las puntuaciones factoriales de las provincias (individuos-compromiso) en los 3 primeros ejes de las exportaciones y en los 3 primeros ejes de las importaciones (6 variables en total). Con ello podremos clasificar las provincias castellano-leonesas según

su comportamiento con respecto al comercio exterior. Aplicaremos un método de clasificación jerárquico (método de Ward) y utilizaremos la distancia euclídea.

La figura 2 recoge el dendograma que resulta de tal análisis. Si se repite el análisis por separado, importaciones por un lado y exportaciones por el otro, los resultados son similares.

Figura 2. Dendograma (método Ward)

INDIVIDUOS	Etiqueta	nº	Distancias entre clusters combinados							
			0	5	10	15	20	25		
	Segovia	6	--+							
	Zamora	9	--+							
	Soria	7	--+							
	Ávila	1	--+							
	León	3	--+-----+							
	Salamanca	5	--+ +-----+							
	Burgos	2	-----+						+-----+	
	Palencia	4	-----+						+-----+	I
	Valladolid	8	-----+						+-----+	

El resultado del cluster permite corroborar las conclusiones obtenidas hasta ahora. Hay un primer grupo de provincias de Castilla y León (formado por Segovia, Zamora, Soria, Ávila, León y Salamanca) que presentan un comportamiento homogéneo respecto al tipo de productos exportados e importados (provincias con poca industria y unidades productivas de pequeña dimensión). Por otro lado aparecen Burgos, Palencia y Valladolid, cada una de ellas con ciertas peculiaridades respecto a las importaciones y a las exportaciones, algunas de las cuales ya se han puesto de manifiesto al interpretar los resultados del análisis STATIS. Estas son precisamente las provincias con más industria y que tradicionalmente con empresas de carácter transnacional.

Sería interesante analizar las provincias de Castilla y León del primer cluster con el fin de caracterizarlas mejor. Este puede ser el objeto de un nuevo estudio.

4. CONCLUSIONES

Con el análisis realizado se ha conseguido el objetivo previsto: establecer diferencias entre las provincias de Castilla y León en lo que al comercio exterior se refiere. Por una lado, encontramos las provincias que marcan la tónica general de la comunidad (Ávila, Zamora, León, Segovia, Soria y Salamanca), y de ahí su posición cercana al origen en la mayoría de los gráficos obtenidos.

Por otro lado, las provincias que presentan un comportamiento diferente son aquellas en las que es muy significativa la presencia de empresas con un elevado

volumen de intercambios con el exterior, tales como GRUPO FASA-RENAULT (Palencia y Valladolid), GLAXO WELLCOME y LOREAL-PROCASA (Burgos) o ENERTEC (Valladolid). Este comportamiento se acentúa con el paso del tiempo puesto que los puntos representativos del año 96 son los más extremos en todos los casos.

Asimismo, se han puesto de manifiesto las posibilidades del método STATIS a la hora de analizar datos económicos observados en diferentes momentos del tiempo, y del análisis cluster como técnica complementaria a la hora de interpretar los resultados.

5. BIBLIOGRAFÍA

- GLAÇON, F.: Analyse conjointe de plusieurs matrices de données. Thèse de troisième cycle, Université Scientifique et Médicale de Grenoble, 1981.
- GROUPE D'ÉTUDE DE RÉFLEXION INTERRÉGIONALE: L'analyse des données évolutives. Méthodes et applications. Éditions Technip. Paris, 1996.
- JUNTA DE CASTILLA Y LEÓN: Anuario Estadístico de Castilla y León. 1997.
- LAVIT, C.: Analyse conjointe de tableaux quantitatifs. Éditions Masson. Paris, 1988.
- LEBART, L.; MORINEAU, A.; PIRON, M.: Statistique exploratoire multidimensionnelle. Dunod. Paris, 1995.
- L'HERMIER DES PLANTES, H.: Structuration des tableaux à trois indices de la statistique. Thèse de troisième cycle. Université de Montpellier, 1976.
- PÉREZ HUGALDE, C.: Evolución de las relaciones entre la productividad del trabajo en la agricultura, las estructuras agrarias y el desarrollo económico en España (1960-1979). Un análisis mediante métodos estadísticos multivariados. Tesis doctoral de tercer ciclo, Universidad Politécnica de Madrid, 1987.