

ANÁLISIS COMPARATIVO DE LA PRODUCTIVIDAD CEREALISTA DE LAS PROVINCIAS DE CASTILLA Y LEÓN DESDE 1991 A 1995 MEDIANTE LA INTEGRACIÓN DE SUBESPACIOS BIPLLOT.

DORADO DÍAZ, Ana VICENTE TAVERA, Santiago
MARTÍN RODRÍGUEZ, Jesús BLÁZQUEZ ZABALLOS, Antonio
Dpto. de Estadística y Matemática Aplicadas.
Universidad de Salamanca.

1.- INTRODUCCIÓN:

La Política Agraria Comunitaria (PAC), ha continuado aplicándose en España con la orientación, ya iniciada en 1992, de conseguir reducir la extensión de tierra dedicada a la producción de algunos productos agrícolas. Fundamentalmente consiste en reducir paulatinamente el sistema de precios de garantía, compensando a los agricultores con subvenciones directas al agricultor en función de su producción y de las tierras que deja de cultivar. El volumen de subvenciones recibidas en 1995 se eleva a 695,8 miles de millones, con un aumento del 18,1% sobre las del año anterior (*BBV: "Informe Económico 95"*).

Pese a la pérdida de importancia relativa que ha sufrido el Sector Agrario frente a la Industria y al Sector Servicios, es indudable que su desarrollo es fundamental para el crecimiento económico de un país. Uno de los indicadores que mejor cuantifica el nivel de desarrollo agrario es la "productividad de la tierra", concepto que mide el cociente entre la producción total y una media ponderada de factores (en nuestro caso superficie agraria cultivada).

Los economistas contemporáneos coinciden en que una mejora en la productividad de la tierra es un paso indispensable para un posterior desarrollo industrial. Conocer pues, un indicador que mida el nivel de desarrollo agrario y su evolución en el tiempo

constituye un tema de gran importancia. Por ello hemos construido una matriz de datos que recoge productividades de cereales, tubérculos etc. en las distintas provincias de la Comunidad de Castilla y León, durante el periodo 1991-1995. Sobre esta matriz analizaremos los posibles cambios de estructura de la productividad agraria a través de un procedimiento estadístico basado en representaciones HJ-Biplot (GALINDO, 1986).

2.- MATERIAL Y MÉTODOS:

Para la realización del estudio, partimos de los datos referentes a producciones (medidas en miles de Tm.) y superficies agrarias (medidas en miles de Ha.) de las provincias de Castilla y León durante los años: 1991, 1992, 1993, 1994 y 1995, publicados en la revista "Coyuntura Económica de Castilla y León".

Hallando el cociente entre ambas variables (producción y superficie) obtenemos un indicador de la productividad de la tierra con el que poder hacer un análisis comparativo de las distintas provincias durante el período señalado.

Las unidades a clasificar en este caso son pues, las diferentes provincias de la comunidad castellano-leonesa; tomadas por orden alfabético:

-
- | | |
|-------------------|--------------------|
| 1. ÁVILA (AV) | 6. SEGOVIA (SG) |
| 2. BURGOS (BU) | 7. SORIA (SO) |
| 3. LEÓN (LE) | 8. VALLADOLID (VA) |
| 4. PALENCIA (PA) | 9. ZAMORA (ZA) |
| 5. SALAMANCA (SA) | |
-

Las variables sobre las que se han tomado los datos son las siguientes:

-
- | | |
|------------------|--------------------|
| 1. AVENA (AVE) | 5. MÁIZ (MAI) |
| 2. CEBADA (CEB) | 6. PATATA (PAT) |
| 3. CENTENO (CEN) | 7. REMOLACHA (REM) |
| 4. GIRASOL (GIR) | 8. TRIGO (TRI) |
-

2.1.- Métodos Biplot:

" Un Biplot es una representación gráfica de una matriz X ($n \times p$) por medio de unos marcadores g_1, \dots, g_n para las filas y unos marcadores h_1, \dots, h_p para sus columnas, elegidos de forma que el producto interno $g_i' h_j$ represente al elemento x_{ij} de la matriz X " (GABRIEL, 1971).

Para una matriz Y de orden $n \times m$, un Biplot puede construirse a partir de la descomposición en valores singulares de la matriz (SVD).

$$Y = UDV'$$

donde U es la matriz de vectores propios de la matriz YY' , D es la matriz de valores propios de la matriz anterior, y V es la matriz de vectores propios de la matriz $Y'Y$.

Hay varias factorizaciones posibles:

- **RMP** se utiliza para aludir al hecho de que con esa factorización se preserva la métrica para las filas (**Row Metric Preserving**).

La representación en "s" dimensiones se hace tomando las "s" primeras columnas de $J=UD$ y $K=V$ como marcadores para filas y columnas respectivamente.

- **CMP** aludiendo al hecho de que con esa factorización se preserva la métrica para las columnas (**Colum Metric Preserving**).

La representación en "s" dimensiones se hace tomando las "s" primeras columnas de $G=U$ y $H=VD$ como marcadores para filas y columnas respectivamente.

Tradicionalmente estos Biplot eran identificados como **JK** y **GH**, pero las referencias más actuales optan por la nueva notación, que fue introducida por GREENACRE (1984).

Ambas representaciones poseen diferentes propiedades. En el GH, se obtiene una alta calidad de representación para las variables, pero muy pobre para las observaciones, mientras que en el JK, se obtiene una alta calidad de representación para las observaciones y muy pobre para las variables. No se puede hablar, entonces, de una representación simultánea en sentido estricto.

Una modificación de los métodos Biplot introducidos por Gabriel es la conocida como representación **HJ-Biplot** (GALINDO, 1986).

2.2.- HJ-Biplot:

GALINDO (1986) utilizando las ideas básicas de los métodos Biplot y del Análisis Factorial de Correspondencias (técnica de representación simultánea que produce una representación geométrica de las filas y las columnas de una tabla de contingencia), propone una nueva técnica que denomina HJ-Biplot, y que utiliza como marcadores para las filas $G=UD$ y como marcadores para las columnas $H=VD$.

Se demuestra que en el HJ-Biplot se consigue la misma calidad de representación para filas que para columnas y que, además, es más alta que en los Biplot clásicos de Gabriel.

Se demuestra también que ambos marcadores pueden representarse en el mismo sistema de referencia ya que cada coordenada para una fila puede expresarse en función de las de las columnas y viceversa.

En principio se interpreta la proximidad entre individuos como similaridad, los ángulos entre dos vectores que unen dos variables con el origen como correlación y la proximidad de un grupo de marcadores fila a un marcador columna en términos de preponderancia. Más concretamente, dos variables separadas por un ángulo pequeño están fuertemente relacionadas, y un grupo de individuos próximo a una variable indica que los individuos próximo a una variable indica que los individuos han tomado valores preponderantes para esta variable.

Proyectando los individuos sobre los vectores que representan a las variables es posible determinar la distribución aproximada que los individuos toman sobre esa variable (*GABRIEL, 1990*).

A pesar de que la interpretación parece sencilla, hay que hacerla con cuidado ya que las posiciones de los puntos sobre los planos principales puede ser sólo aparente. Para evitar los posibles errores de interpretación, tenemos en cuenta algunas de las medidas propuestas por *GALINDO* y *CUADRAS (1986)* para la interpretación de los gráficos resultantes de un HJ-Biplot.

*Contribuciones del elemento al factor: son las partes tomadas por cada elemento en la varianza explicada por el factor.

*Contribuciones del factor al elemento: parte de la dispersión de un elemento explicada por el factor.

*Calidad de representación: parte de la información de un elemento contenida en un eje, plano, etc.

3.- RESULTADOS:

3.1.- Análisis individualizado año a año:

El estudio que se ha llevado a cabo, ha consistido en el análisis individualizado de las representaciones HJ-Biplot de las productividades de las provincias de la Comunidad de Castilla y León año a año (de 1991 a 1995), estandarizando por columnas.

La absorción de inercia para todos los años del estudio en los tres primeros ejes, así como la inercia acumulada ha sido la siguiente:

ANO	EJE 1	EJE 2	EJE 3	I. ACUM.
1991	52.6823	23.1442	9.2119	85.0385
1992	62.9839	17.0490	9.7280	89.7610
1993	45.0598	22.0078	13.8223	80.8899
1994	51.1765	18.5106	17.3377	87.0248
1995	45.1780	21.0091	13.5065	79.6935

Tabla 1: Absorción de inercia de las representaciones HJ-Biplot desde 1991 a 1995

Como podemos apreciar en la Tabla 1, la absorción de inercia acumulada en los tres primeros ejes es muy similar durante los cinco años del estudio, aunque la distribución por ejes varía, ya que por ejemplo en el primer eje el año 1992 tiene una absorción de inercia mucho mayor que el resto de los años, mientras que en los ejes segundo y tercero la variabilidad de la absorción de inercia de unos años a otros es relativamente pequeña. Por lo tanto, las estructuras en relación a las inercias acumuladas se mantienen durante el período, oscilando levemente de un 79,693% a un 89,761% en los tres primeros ejes.

Los análisis efectuados nos han permitido observar como las estructuras para las variables (cereales, tubérculos, etc.) mantienen cierto parecido entre los años 1991-1993 y 1992-1994, quedando 1995 como un año de cambio con respecto a los anteriores. El hecho de que el año 1995 difiera del resto, puede deberse a que en años anteriores la productividad total se adjudicaba en términos subjetivos, por lo tanto las estadísticas publicadas durante este tiempo no son muy fiables. A finales de 1994, el agricultor empieza a declarar la superficie cultivada así como la producción para poder beneficiarse de las ayudas de la PAC, por lo que los datos a partir de ese momento se acercan más a los reales.

Al analizar en el año 1991 el primer plano principal (ver Fig.1), observamos como el eje 1 separa la variable maíz (MAI) del resto de variables generalmente de secano como son: la cebada (CEB), el trigo (TRI), la avena (AVE) y el centeno (CEN).

El eje 2 separa la variable girasol (GIR), y maíz en menor medida, de la remolacha (REM).

La máxima calidad de representación para todas las variables (productos cultivados) está en el plano 1-2, a excepción de la patata (PAT), que tiene su contribución más alta en el eje 3 (ver Tabla 2). En el plano 1-3 la patata separa a Valladolid del resto de provincias .

VARIABLES	EJE 1	EJE 2	EJE 3
TRI	807.6288	71.2100	14.4157
CEB	894.5134	0.8817	34.9541
AVE	819.7760	24.9836	4.6738
CEN	660.6646	53.0710	109.5747
MAI	405.6210	220.0505	79.2881
PAT	451.7189	23.4378	488.9523
REM	28.0041	772.6331	2.1157
GIR	146.6603	685.2704	2.9790

Tabla 2: Contribuciones relativas del factor al elemento (columnas)

Este posicionamiento de variables en el primer plano principal, nos permite diferenciar a las provincias de Salamanca y Ávila, como provincias cuya rentabilidad en el maíz es más alta, frente a provincias como Burgos y Palencia, cuya rentabilidad máxima se da en los cereales de secano. El eje 2 nos separa Segovia por su gran productividad en remolacha frente a León por el girasol. El resto de provincias en este plano tienen baja calidad de representación .

La estructura analizada en el año 1991 es muy similar a la de 1993, especialmente en la distribución de las variables. En el plano principal 1-2 (ver Fig. 2) observamos cómo la remolacha y el maíz, que tienen más contribución en el eje 2 que en el eje 1, se oponen al trigo, cebada, avena y centeno. Por contra, la estructura para las provincias ha variado respecto a 1991; ahora el eje 1 separa Zamora de Burgos por la alta rentabilidad de ésta en cereales de secano, y el eje 2 separa Ávila y Palencia (por su alta productividad en remolacha) por un lado, de Valladolid y León (por su rentabilidad en maíz), y por otro de Soria por su alta productividad en patata (ésto queda mejor representado, en el plano 2-3, puesto que las contribuciones de Soria y de la patata son máximas en el eje 3.

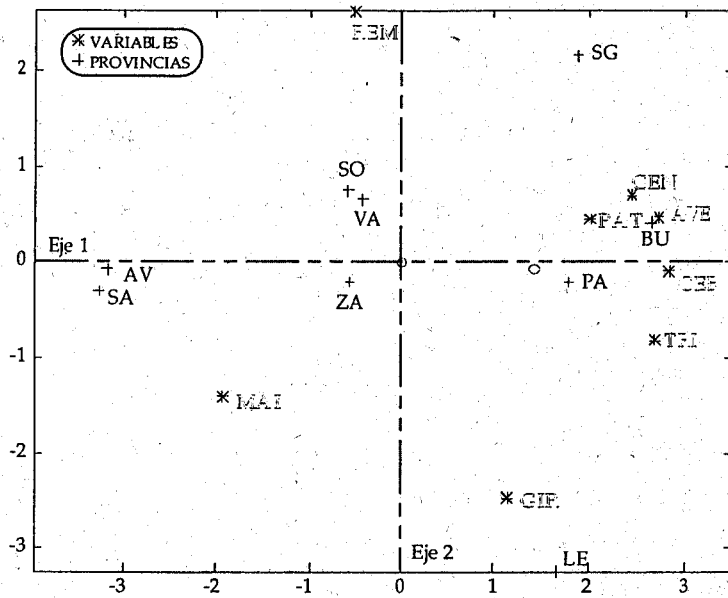


Fig. 1: Representación HJ-Biplot de las productividades en Castilla y León en el plano 1-2 (Año 1991)

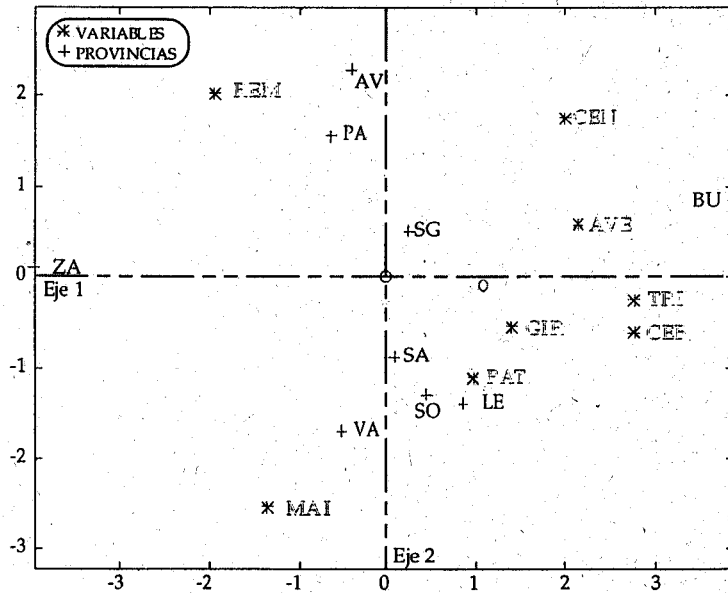


Fig. 2: Representación HJ-Biplot de las productividades en Castilla y León en el plano 1-2 (Año 1993).

El análisis del tercer eje (ver Fig. 3) en los años 1991 y 1993 es muy similar para las variables puesto que la patata es la que mayor contribución aporta al mismo. Vemos que en 1991 Valladolid se caracteriza por tener una alta rentabilidad en patata que la diferencia del resto de provincias, especialmente de Soria, mientras que en 1993 se invierte el proceso anterior y es Soria la de mayor rentabilidad en patata.

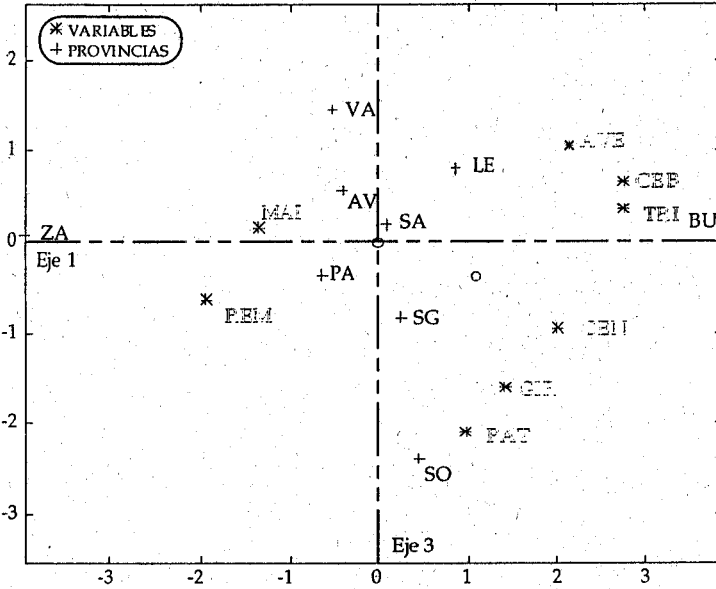


Fig. 3: Representación HJ-Biplot de las productividades en Castilla y León en el plano 1-3 (Año 1993).

Las estructuras de las variables en los planos principales de los años 1992 y 1994 también son bastante similares. En el plano 1-2 de 1992 (ver Fig. 4) observamos como el primer eje separa la remolacha de los cereales de secano, mientras que el segundo eje separa las patatas del maíz. La distribución del primer plano principal de 1994 es prácticamente igual con la excepción del girasol, que adquiere mayor importancia en el eje 2, aunque se siga manteniendo la posición de la patata y el maíz (ver Fig. 5).

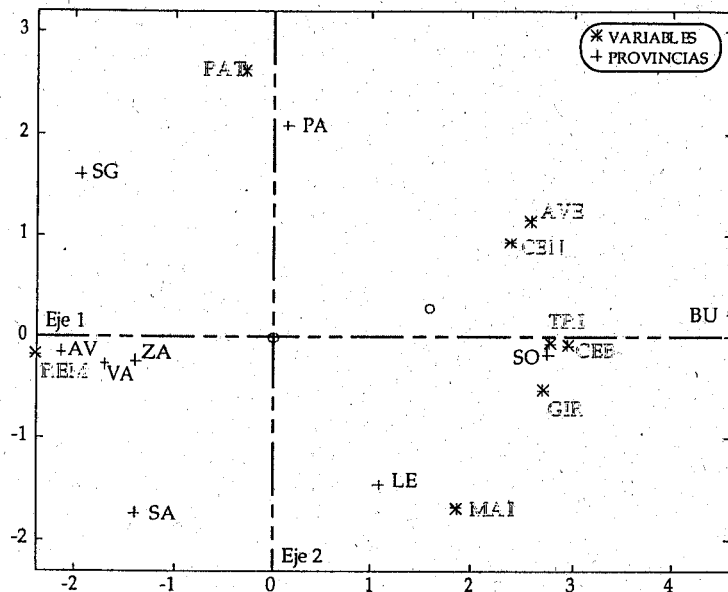


Fig.4: Representación HJ-Biplot de las productividades en Castilla y León en el plano 1-2 (Año 1992).

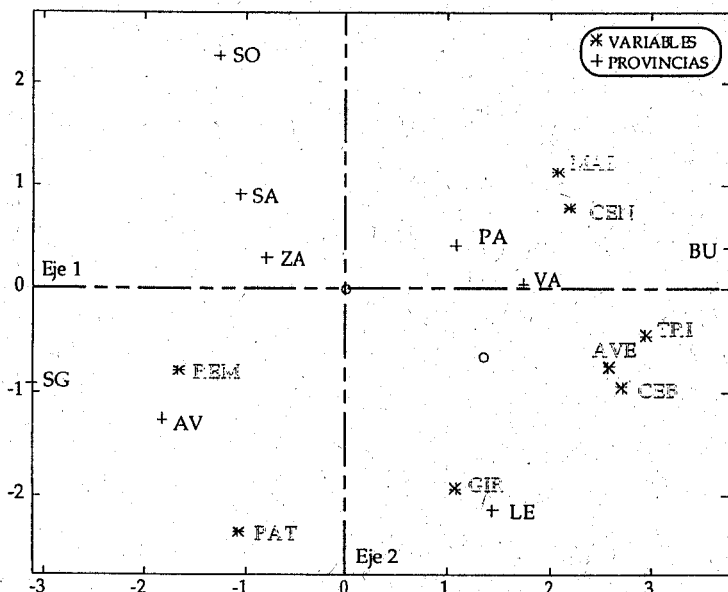


Fig.5: Representación HJ-Biplot de las productividades en Castilla y León en el plano 1-2 (Año 1994).

La estructura de provincias de nuevo es diferente en estos dos años puesto que el primer eje en 1992 (ver Fig. 4) separa las provincias de Ávila, Valladolid y Zamora (por su importancia en la remolacha) de Burgos y en menor medida de Soria, por la importancia de éstas en los cereales de secano. El segundo eje separa León, por su importancia en el maíz, de Palencia por su importancia en la patata, quedando Salamanca y Segovia en posiciones opuestas e intermedias (interpretación de plano), influenciadas Salamanca por el maíz y la remolacha y Segovia por la patata y la remolacha. En 1994 el primer eje separa Segovia, y en menor medida Ávila, por su rentabilidad en remolacha, de Burgos, por los cereales de secano. Por otro lado, el eje 2 separa León, por su importancia en el girasol, de Soria y en menor medida de Salamanca por su bajo rendimiento en patata y girasol.

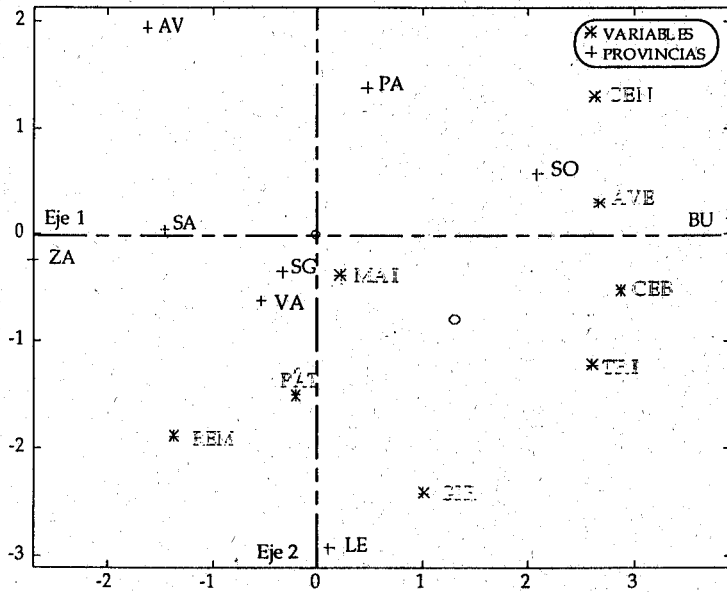


Fig. 6: Representación HJ-Biplot de las productividades en Castilla y León en el plano 1-2 (Año 1995).

La estructura para las variables en el año 1995 parece dar a entender un cambio con respecto a los años anteriores, aunque se mantienen ciertas similitudes entre todos ellos. Así el primer eje (ver Fig. 6) separa la remolacha de los cereales de secano, aunque estos últimos ahora tienen mayor dispersión, y por lo tanto, mayor incidencia en el segundo eje. El eje 2 opone las variables girasol y patata del centeno principalmente. El maíz tiene una calidad de representación en este plano muy baja, ya que es una variable eminentemente de eje 3 (ver Fig. 7).

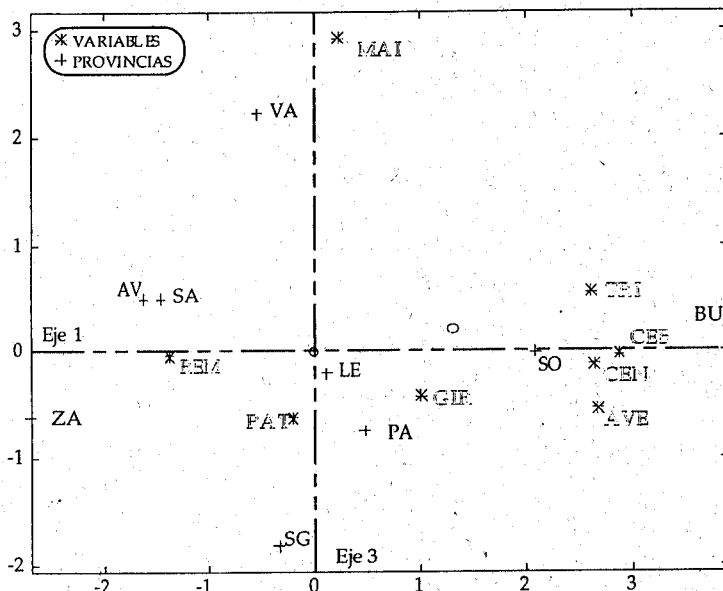


Fig.7: Representación HJ-Biplot de las productividades en Castilla y León en el plano 1-3 (Año 1995).

La estructura de las provincias (ver Fig.6 y 7) es más variable que en los años anteriores. Así, en el primer plano, vemos como Zamora y Salamanca se oponen, por su mayor rendimiento en remolacha, a Burgos y Soria, de nuevo por su alta productividad en los cereales de secano. El segundo eje separa León de Ávila y Palencia, por la importancia de León en girasol, y en menor medida en remolacha y patata, frente a los bajos rendimientos de Ávila y Palencia en éstos productos. A su vez Palencia se separa de Ávila por tener mayor rentabilidad en centeno. Segovia y Valladolid tienen baja calidad de representación en ese plano, porque son provincias eminentemente de eje 3 (ver Tabla 3). En el plano 1-3, observamos como Valladolid adquiere gran rentabilidad en maíz durante este año, por lo que se opone a Segovia.

PROVINCIAS	EJE 1	EJE 2	EJE 3
AV	321.0221	462.3387	29.6420
BU	949.8245	1.3898	2.8236
LE	1.4506	854.3017	4.1977
PA	38.4815	309.5843	91.3786
SA	595.6632	1.0628	69.8236
SG	23.4791	24.4016	695.9635
SO	570.9581	43.3704	0.0349
VA	44.0063	56.5992	739.9031
ZA	828.6252	7.0550	44.9296

Tabla 3: Contribuciones relativas del factor al elemento (filas)

3.2.- Análisis conjunto del período 1991-1995:

En el siguiente estudio, realizamos un análisis conjunto de la representación conjunta HJ-Biplot de las productividades de las provincias de Castilla y León de 1991 a 1995 (ver Fig. 8). El procedimiento que se plantea, es similar al propuesto por KRZANOWSKI, W.J. (1979) y posteriormente ampliado por MARTÍN-RODRÍGUEZ, J. (1996), pero no intentamos la búsqueda de un espacio conjunto, sino que integramos todo el análisis conjuntamente mediante una matriz de 45 filas (provincias castellano-leonesas en los distintos años), por 8 columnas (productos cultivados utilizados en el estudio anterior). Las abreviaturas utilizadas tanto para provincias como para variables son las mismas del estudio individualizado. El número que acompaña a cada abreviatura de provincia, hace referencia a: "1" para el año 1991, "2" para el año 1992, "3" para 1993, "4" para 1994 y "5" para 1995.

La inercia acumulada en los tres primeros ejes principales es la siguiente:

EJE 1	EJE 2	EJE 3
43.2975	60.6803	74.4562

Tabla 4: Inercia acumulada de la representación conjunta.

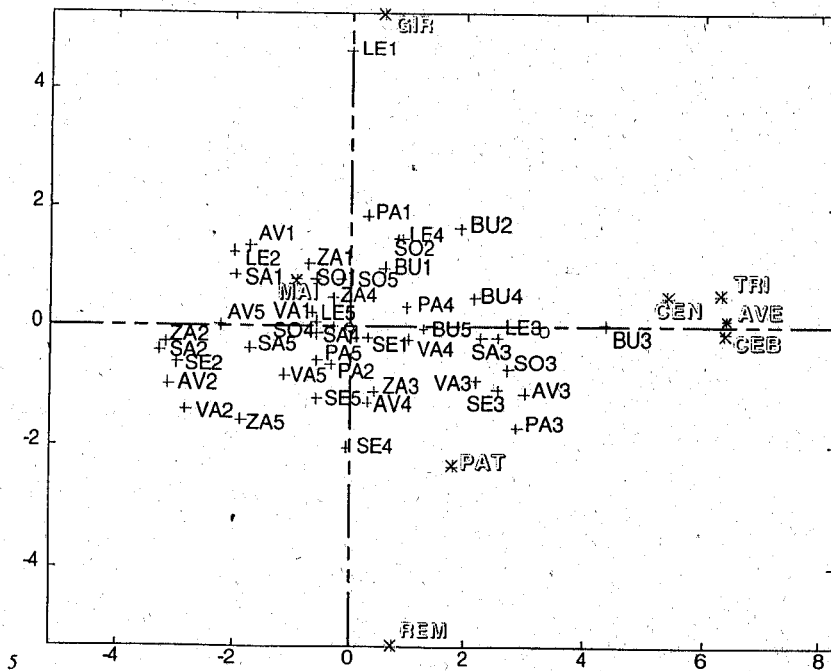


Fig.8: Representación conjunta HJ-Biplot de las productividades en Castilla y León en el plano 1-2 durante el período de 1991 a 1995.

Para interpretar el primer plano principal del análisis conjunto en cada uno de los años analizados, eliminamos de la representación conjunta las provincias correspondientes a los otros cuatro años (ver Fig. 9).

Analizamos en primer lugar la estructura para las variables en la representación conjunta, y vemos que el eje 1 separa los cereales de secano del maíz, y el eje 2, el girasol de la remolacha y la patata.

En el año 1991, cabe destacar la gran rentabilidad de León en girasol. Palencia y Burgos son las más influenciadas por su productividad en cereales de secano, mientras que el resto de provincias (excepto Segovia, que tiene una calidad de representación baja en este plano) tienen una rentabilidad mayor en maíz.

En el año 1992, León pierde rentabilidad en girasol y la adquiere en maíz. Burgos y Soria son las más rentables en cereales de secano y el resto de provincias se ven influenciadas tanto por el maíz como por las patatas (Palencia es la provincia con menor calidad de representación en este plano).

En el año 1993 Burgos nuevamente es la provincia con mayor rentabilidad en cereales de secano. Como podemos observar, parece que la remolacha así como el maíz y los cereales, adquieren mayor rentabilidad para todas las provincias de la Comunidad durante este año y a su vez, la productividad del girasol y del maíz ha disminuido.

Durante el 1994, la productividad del girasol en León vuelve a aumentar. Burgos sigue siendo la más rentable en la producción de cereales de secano, y Segovia, Ávila y Valladolid son las más rentables en patatas y en menor medida en remolacha. La importancia del maíz ha aumentado con respecto a 1993.

En el 1995*, vuelve a ser el maíz el producto más rentable para la mayor parte de las provincias. Burgos sigue siendo la provincia con mayor productividad en cereales de secano y Segovia y Zamora en patata y, en menor medida, en remolacha.

La estructura de las variables en la representación conjunta, vemos que es bastante parecida a la que se daba en el estudio anterior en el año 1991. El eje 1, separa maíz de los cereales de secano y el eje 2, el girasol de la remolacha y la patata.

Aunque la estructura más parecida sea la de 1991, vemos que se cumplen pautas comunes en todos los años.

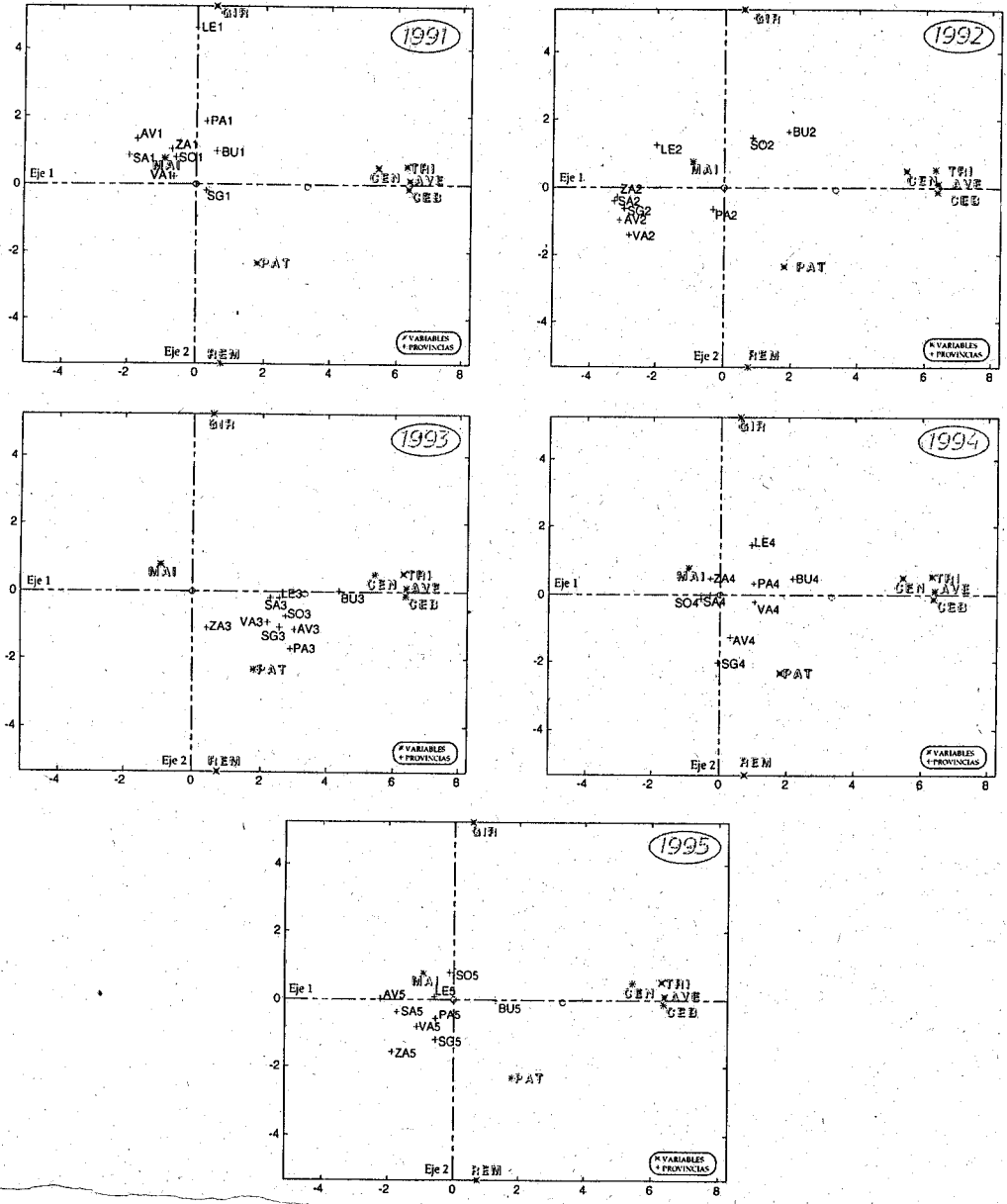


Fig. 9: Análisis individualizado de la representación consenso HJ-Biplot de las productividades de las provincias de Castilla y León de 1991 a 1995 (sacado de la Fig. 8).

Los cereales de secano siempre aparecen juntos, sin oponerse unos a otros, esto es debido a que la superficie dedicada a su cultivo varía poco. A su vez, vemos que es siempre la provincia de Burgos la que obtiene mayor rentabilidad en estos productos, debido al suelo y clima propicios.

Es importante apreciar la pérdida de importancia relativa de la productividad del girasol, posiblemente ocasionada por dos motivos fundamentales: la disminución gradual en la cuantía de las subvenciones, y que el cultivo del girasol deteriora bastante el terreno. Por otra parte, las ayudas a la retirada de tierras y el cultivo de productos no comestibles han aumentado, lo que hace que a veces sea más rentable dejar la tierra en barbecho, que dedicarla al cultivo del girasol.

El cultivo de la patata también presenta características específicas. No hay unos ciclos claramente establecidos, si el agricultor cree que al empezar el año el cultivo de patata puede ser bueno, siembra y si no es así, siembra maíz.

Al estudiar la estructura de las provincias en la representación conjunta para los cinco años, vemos que es muy variable. En los gráficos podemos observar un cierto comportamiento cíclico en las provincias. En un principio, el girasol y el maíz son los productos más rentables para la mayoría de provincias. En 1992, el girasol pierde importancia y la van adquiriendo la patata y la remolacha. Al pasar a 1993, adquieren mucha importancia los cereales de secano y la patata en detrimento del maíz. En 1994 parece que el girasol vuelve a adquirir rentabilidad y en 1995 vuelve otra vez al maíz.

4.- CONCLUSIONES:

1ª Se comprueba que mediante el análisis individualizado año por año, podemos ver que las estructuras para las variables son bastante semejantes a lo largo de los cinco años de estudio.

2ª El análisis conjunto, al compararse con los análisis individualizados, nos permite ver cómo evolucionan las productividades agrarias a lo largo del periodo analizado.

3ª Por último, podemos concluir, que Castilla-León tiene una productividad más o menos constante a lo largo de estos años, aunque la distribución de los cultivos por provincias se vaya alterando. Según expertos agrícolas, la causa más importante para esta alternancia en los cultivos es el clima.

5.- BIBLIOGRAFÍA:

GABRIEL, K. R. (1971). The biplot-graphic display of matrices with application to principal component analysis. *Biometrika*, 58, pp.453-467.

GABRIEL, K. R. (1981). Biplot. *Encyclopedia of Statistical Sciences*. S. Koltz & N. L. Jonhson. Wiley. New York.

GABRIEL, K. R y ODOROFF, C. L. (1990). Biplot in biomedical research. *Statistics in Medicine*, 9, pp. 469-485.

GALINDO, M. P. (1986). Una alternativa de representación simultánea: HJ-Biplot. *Questío*, 10(1), pp. 13-23.

GALINDO, M. P. y CUADRAS, C. M. (1986). Una extensión del método Biplot a su relación con otras técnicas. *Publicación de Bioestadística y Biomatemática*. Universidad de Barcelona. N° 17.

INFORME ECONÓMICO 95. Servicios generales del Banco Bilbao-Vizcaya. Bilbao, 1996.

KRZANOWSKI, W.J. (1979). Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74 (367), pp. 703-707. Corrección en 76: 1022.

MARTÍN-JESÚS, J. (1996). Contribuciones a la Integración de Subespacios desde una perspectiva Biplot. Tesis Doctoral. Universidad de Salamanca.