6. HJ-Biplot como herramienta de inspección de matrices de datos bibliométricos

Artículo publicado en la Revista Española de Documentación Científica 36(1):e001. doi: http://dx.doi.org/10.3989/redc.2013.1.988. *Autores: Adrián A. Díaz-Faes, Borja González-Albo, Mª Purificación Galindo, & María Bordons.*

Resumen: El objetivo de este trabajo es poner de manifiesto la utilidad del HJ-Biplot en los estudios bibliométricos. Una representación intuitiva y sencilla, similar a un diagrama de dispersión, pero que captura las estructuras de covariación multivariantes entre los indicadores bibliométricos. Su interpretación no requiere conocimientos estadísticos especializados, basta con saber interpretar la longitud de un vector, el ángulo entre dos vectores y la distancia entre dos puntos. Con este fin, se analiza la actividad científica de los centros propios y mixtos del CSIC durante el período 2006-2009 mediante una serie de indicadores de colaboración e impacto científico. Utilizando un HJ-Biplot es posible interpretar simultáneamente la posición de los centros, representados por puntos, de los indicadores, representados mediante vectores, y de las relaciones entre ambos, en el plano con mayor capacidad informativa. Esto nos permite analizar la situación de cada centro en el contexto de su área y en el contexto general del CSIC e identificar aquéllos que muestran un comportamiento singular. Se concluye que las áreas de Humanidades y Ciencias Sociales, Biología y Biomedicina, Materiales y Químicas son más homogéneas en el comportamiento de sus centros, mientras que Físicas, Agrarias, Recursos Naturales y Alimentos muestran mayor heterogeneidad.

Palabras clave: HJ-Biplot, análisis multivariante, bibliometría, colaboración científica, CSIC.

6.1. Introducción y objetivos

El término “Bibliometría” se atribuye a Pritchard (1969), quien definió el campo como “la aplicación de métodos matemáticos y estadísticos a los libros y otros medios de comunicación”. Años más tarde, Subramanyam (1983) señala que el método bibliométrico facilita el estudio de las relaciones entre las variables a través de la aplicación de técnicas estadísticas como la regresión, correlación o análisis factorial. En las últimas décadas hemos asistido a un importante desarrollo de los estudios bibliométricos, que incluyen indicadores cada vez más sofisticados, un creciente uso de técnicas estadísticas y el desarrollo de nuevas técnicas de visualización de la información. Aunque en los últimos años ha sido especialmente importante el auge

que han experimentado las técnicas de visualización, muy por encima del observado para las técnicas de análisis multivariante, estas últimas también constituyen una interesante herramienta en bibliometría, y a ellas nos vamos a referir en este artículo.
En los estudios bibliométricos abundan representaciones descriptivas uni y bivariantes, siendo las técnicas multivariantes más utilizadas el Análisis de Cluster, el Análisis Factorial con solución en Componentes Principales, y el Análisis de Correspondencias.
El Análisis de Cluster permite clasificar las unidades según similitud, pero no es posible saber qué combinación de variables es la que motiva los agrupamientos que exhibe el correspondiente dendograma. El Análisis Factorial está encaminado a buscar unas pocas variables hipotéticas (conocidas como factores o variables latentes), generadas a partir de las variables observables, que capturen la mayor parte de la información contenida en los datos, pero no proporciona información sobre la similitud entre las unidades objeto de estudio.

La utilización de métodos de representación simultánea de datos multidimensionales se ha visto reducida al Análisis de Correspondencias (Benzécri, 1973), técnica íntimamente relacionada con el Análisis de Componentes Principales, que permite visualizar la posible relación entre un par de variables categóricas, y entre sus respectivas categorías, pero está pensada para trabajar con matrices de frecuencias. En el ámbito bibliométrico esta técnica ha sido empleada por el CNRS para mostrar la evolución de los patrones de publicación a lo largo del tiempo (Miquel y otros, 1995; Doré y otros, 1996; Okubo y otros, 1998; Doré y Ojasoo, 2001), en el análisis jerárquico de la coautoría en las redes de colaboración (Abd el Kader y otros, 1998), así como en el análisis de patentes (Doré y otros, 2000). El Análisis de Correspondencias también ha sido aplicado por Bordons y otros (2004) para estudiar las tendencias en la investigación sobre la aspirina, por Sanz-Casado y Conforti (2005) para analizar la relación entre tipologías documentales y pautas de colaboración científica a nivel micro, por Anuradha y Urs (2007) para identificar patrones de colaboración entre investigadores de la India y por Nagpaul (1995) y Súarez-Balseiro y otros (2009) para evaluar la contribución de los investigadores de las universidades de la India y Puerto Rico, respectivamente, a las publicaciones científicas de mayor impacto internacional.
El HJ-Biplot propuesto originalmente por Galindo (1986), presenta las ventajas del Análisis de Correspondencias, pero es aplicable a cualquier matriz de datos, no solo frecuencias. A pesar de ser una técnica para inspección de matrices de datos multivariantes con menos restricciones que las Correspondencias o el Análisis Factorial, la única referencia de su utilización en el ámbito bibliométrico es Díaz-Faes y otros (2011), donde se aplica el HJ-Biplot para analizar la actividad científica de un conjunto de universidades en el área biosanitaria. No ocurre así en otros campos de la ciencia. Ver, por ejemplo, Cárdenas y otros (2007) que citan aplicaciones en medicina, economía, biología o tecnología ambiental entre otras. Una referencia actual y particularmente interesante es Caballero-Juliá (2011), que presenta el HJ-Biplot como

herramienta en el análisis de grupos de discusión y lo aplica a datos de calidad de vida en ludopatía, en la que puede consultarse exhaustivamente el método.

Para poner de manifiesto la utilidad de la técnica en los estudios bibliométricos, en este estudio se caracteriza la producción científica de los centros propios y mixtos del CSIC en el período 2006-2009 en relación a una serie de indicadores bibliométricos de impacto y colaboración científica. La influencia de la colaboración sobre el impacto de la producción de la investigación de los centros del CSIC ha sido objeto de análisis en un estudio previo (González-Albo y otros, 2012), pero en este caso se presenta una aproximación multivariante a partir de un análisis integrado de indicadores. Se pretende mostrar la utilidad del HJ-Biplot para analizar la relación entre las variables estudiadas, identificar centros que tienen un comportamiento similar en relación a dichas variables y explorar el grado de homogeneidad de las áreas científico-tecnológicas del CSIC en lo que al comportamiento de sus centros se refiere; todo ello a través de un tratamiento y una representación simultánea de variables y centros en un subespacio de baja dimensión.

6.2. Material y métodos

6.2.1. Objeto de estudio

Se ha trabajado con las publicaciones científicas de los centros propios y mixtos del CSIC, recogidas en la base de datos *Web of Science* (WoS), que incluye el *Science Citation Index Expanded* (SCIE), el *Social Sciences Citation Index* (SSCI) y el *Arts & Humanities Citation Index* (AHCI), durante el período 2006-2009. La identificación y codificación de los centros del CSIC se realizó de forma semi-automática (Morillo y otros, 2013) a partir de la producción científica de España – “Spain” en el campo “Address” – descargada de la base de datos WoS en febrero de 2011. Se asignó a cada instituto o centro un código alfa-numérico que permite el posterior tratamiento automático de los datos y caracterizar la actividad científica de los centros propios y mixtos del CSIC con un alto grado fiabilidad (Gómez y otros, 2011a). Asimismo, se realizó una normalización de los títulos de revistas en función de los diferentes campos identificativos de las mismas en WoS – “Full Journal”, “Abbreviated Journal”, “Serie” e “ISSN” – para su posterior vinculación con los datos de factor de impacto publicados en el *Journal Citation Reports*. El estudio se limita a los ítems citables, que incluyen artículos originales, notas y revisiones. En adelante nos referimos a los ítems citables como artículos. Se ha caracterizado la actividad científica de cada uno de los centros propios y mixtos del CSIC a través de los siguientes indicadores:

a) Indicadores de impacto:

- ✓ *Porcentaje de artículos en el primer cuartil* (Q1), que considera el porcentaje de artículos publicados por cada centro en el 25% de revistas con mayor factor de impacto de cada disciplina (en el caso de revistas asignadas a más de una disciplina se selecciona aquélla en la que ocupa una mejor posición).

- ✓ *Posición normalizada media* (PN), calculada como el cociente entre la posición que ocupa una revista entre las de su disciplina en orden descendente según su factor de impacto y el número total de revistas de la disciplina. Dicho valor se resta de la unidad, de forma que la PN oscila entre 0 y 1. Valores altos de PN indican una buena situación de la revista dentro de su disciplina (Bordons y Barrigón, 1992). Se ha asignado a cada artículo la PN de su revista de publicación, calculándose luego la PN de un centro como el promedio de la PN de todos sus artículos. Se utiliza la PN mejor en el caso de que una revista esté asignada a más de una disciplina. La premisa que subyace al uso del Q1 y la PN es que el factor de impacto de las revistas es un indicador de su prestigio en sus campos de especialización (Moed, 2005). Dado que los valores de factor de impacto varían de forma importante según las disciplinas, se ha preferido utilizar los indicadores Q1 y PN, que no tienen esta limitación y permiten realizar comparaciones entre disciplinas.

- ✓ *Citas relativas al mundo en el período 2006-2009* (CRM), que considera las citas recibidas por los documentos desde el año de publicación hasta 2010 normalizadas respecto a las citas medias recibidas por la producción mundial en cada disciplina. Los datos del total mundial utilizados como referencia proceden de *Thomson Reuters*, que proporcionó una tabla que incluía para cada disciplina las citas medias por artículo recibidas por los artículos de cada año (desde 2005 hasta 2009)[16]. Una explicación detallada sobre la obtención del indicador CRM puede encontrarse en Gómez y otros (2011a). Para el cálculo del indicador citas relativas de cada centro se comparan las citas recibidas por sus artículos con las recibidas por el promedio del mundo, atendiendo a su disciplina y año de publicación. En el caso de revistas asignadas a más de una disciplina en WoS, se han calculado los valores de CRM respecto a las distintas disciplinas, obteniéndose luego el valor medio. Un valor de citas relativas superior a la unidad indica que el centro recibe más citas que el promedio del mundo en sus disciplinas de publicación. El uso de las citas como indicador de la influencia o impacto de la investigación sobre la comunidad científica está ampliamente aceptado en la actualidad, aunque es necesario tener en cuenta sus inconvenientes y limitaciones, repetidamente recogidas en la literatura (Moed, 2005).

[16] Cedidos al MINECO para la convocatoria del Subprograma de Apoyo a Centros y Unidades de Excelencia Severo Ochoa 2011.

b) Indicadores de colaboración:

- ✓ *Porcentaje de artículos sin colaboración* (firmados por un solo centro), *porcentaje de artículos en colaboración nacional* (dos o más centros españoles) y *porcentaje de artículos en colaboración internacional* (al menos un centro extranjero). Los artículos que presentan colaboración nacional e internacional simultáneamente se han considerado en la categoría "internacional".

Hay que señalar que WoS no calcula el factor de impacto para las revistas de Humanidades, por lo que los indicadores Q1 y PN de los centros de estas áreas se refieren sólo a la fracción de su producción que cuenta con dichos indicadores. Por otro lado, se han eliminado del estudio cuatro institutos, adscritos a Humanidades y Ciencias Sociales, por no contar con artículos en revistas con factor de impacto, lo que impedía el cálculo de las variables porcentaje de artículos Q1 y PN: Instituto de Lengua, Literatura y Antropología, Escuela de Estudios Hispanoamericanos, Escuela Española de Historia y Arqueología en Roma e Instituto de Estudios Islámicos y del Oriente Próximo. El conjunto final de datos analizados consistió en una matriz X_{136x6}, 136 centros del CSIC y 6 indicadores bibliométricos.

6.2.2. Métodos Biplot

Los métodos Biplot fueron propuestos por Gabriel (1971) como representaciones gráficas de datos multivariantes, es decir, al igual que un diagrama de dispersión muestra la distribución conjunta de dos variables, un Biplot representa tres o más variables (Gabriel y Odoroff, 1990); son pues, técnicas multivariantes. Usualmente, las filas de la matriz son representadas mediante puntos (marcadores fila) y las columnas con vectores (marcadores columna), siguiendo la terminología introducida por el autor.

Formalmente se definen de la siguiente manera: un Biplot para una matriz de datos X_{nxp}(arreglo rectangular con n filas y p columnas) es una representación gráfica mediante marcadores $g_1, g_2, \ldots., g_n$ para las filas de la matriz de datos X y $h_1, h_2, \ldots., h_p$ para las columnas de X, de forma que el producto escalar $g_i T\, h_j$ aproxime el elemento x_{ij} de la matriz de partida, tan bien como sea posible (Gabriel, 1971). El producto escalar, en el que se basa, es un concepto matemático que en un principio podría suponer una barrera para el usuario, pero su traducción geométrica es sencilla. En este trabajo los datos están contenidos en una matriz X_{136x6} que tiene en filas los 136 centros del CSIC y en columnas, los 6 indicadores bibliométricos. Así, para cada fila i (cada centro del CSIC en nuestro caso) y cada columna j (indicadores bibliométricos) aparece en la matriz de datos un valor x_{ij} que es el valor de ese

marcador j para el centro i. Un Biplot permite representar la fila i de la matriz de datos (un centro del CSIC) con el marcador g_i y la columna j con el vector h_j, de forma que al proyectar el punto g_i sobre el vector h_j, esa proyección coincide con el valor que ese centro ha tenido para ese indicador. Esa es la traducción geométrica del concepto de producto escalar.

El interés práctico reside en que el orden de las proyecciones de cada marcador fila sobre un marcador columna reproduce el orden de la matriz de partida, de forma que analizando la posición de cada unidad (centro) sobre cada variable (indicador), es posible ordenar las unidades en función del valor que toman en ese indicador, y eso puede hacerse para todos y cada uno de los indicadores (ver Figura 6.1).

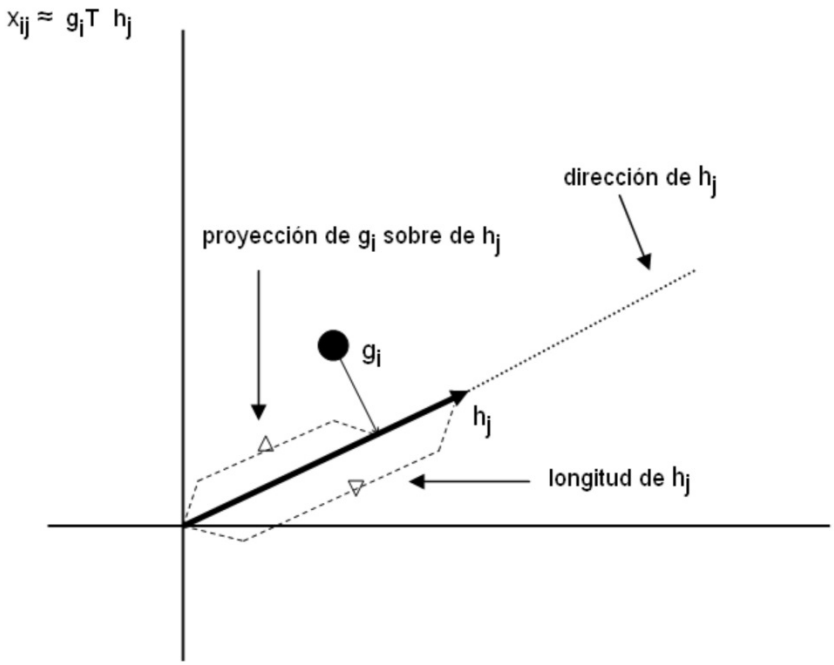

Figura 6.1. Representación geométrica del producto escalar.

Hay infinitas formas de representar un Biplot pero solo algunas tienen propiedades interesantes en el análisis de datos. Los Biplots propuestos originalmente por Gabriel (1971) fueron dos: JK-Biplot, en el cual las filas son representadas con la máxima calidad de representación (fiabilidad de las posiciones de los puntos que representan a los centros) y GH-Biplot en el cual las columnas son representadas con la máxima calidad, pero no las filas.

Galindo (1986) propone el HJ-Biplot como una representación gráfica multivariante de los datos de una matriz X_{nxp}, mediante marcadores $j_1, \dots, j_n$ para las filas y $h_1, \dots, h_p$ para las columnas, elegidos de forma que ambos marcadores puedan ser superpuestos en un mismo sistema de referencia con máxima calidad de representación. Al presentar filas y columnas idéntica bondad de ajuste es posible interpretar no sólo la

posición de las filas y de las columnas, sino también las relaciones fila-columna. Los ejes que conforman el sistema de referencia son las Componentes Principales del espacio de los indicadores.

Las reglas para la interpretación del HJ-Biplot son una combinación de las reglas empleadas en otras técnicas como el Escalamiento Multidimensional, el Análisis de Correspondencias, el Análisis Factorial y los Biplots clásicos (Galindo y Cuadras, 1986). En la Figura 6.2 se muestra un ejemplo con cuatro variables y cuatro centros.

- ✓ Las distancias entre los marcadores fila se interpretan como una función inversa de sus similaridades, de tal forma que marcadores próximos (centros del CSIC) son más similares. Esta propiedad permite la identificación de centros con perfiles similares. Cualquier técnica de agrupamiento jerárquico o no jerárquico se puede utilizar para detectar grupos relevantes (Vicente-Tavera, 1992).

- ✓ La longitud de los marcadores columna (vectores) aproximan la desviación típica de los indicadores bibliométricos.

- ✓ Los cosenos de los ángulos entre los vectores columna aproximan las correlaciones entre los indicadores, de modo que ángulos agudos se asocian a indicadores con alta correlación positiva (variables 1 y 2), ángulos obtusos indican correlación negativa (variables 1 y 4) y ángulos rectos señalan variables no correlacionadas (variables 1 y 3). De la misma manera, los cosenos de los ángulos entre los marcadores de los indicadores y los ejes (Componentes Principales) aproximan las correlaciones entre ambos. Para datos estandarizados, las cargas se aproximan a las de los factores en el Análisis Factorial.

- ✓ El orden de las proyecciones ortogonales de los marcadores fila (puntos) sobre un marcador columna (vector) aproxima el orden de los elementos fila (centros) en esa columna (la misma propiedad se cumple para la proyección de los marcadores columna en la dirección definida por un marcador fila). Cuanto mayor es la proyección de un punto sobre un vector, más se desvía el centro de la media de ese indicador bibliométrico. Para una interpretación correcta, la proporción entre las escalas físicas horizontales y verticales ha de ser la misma.

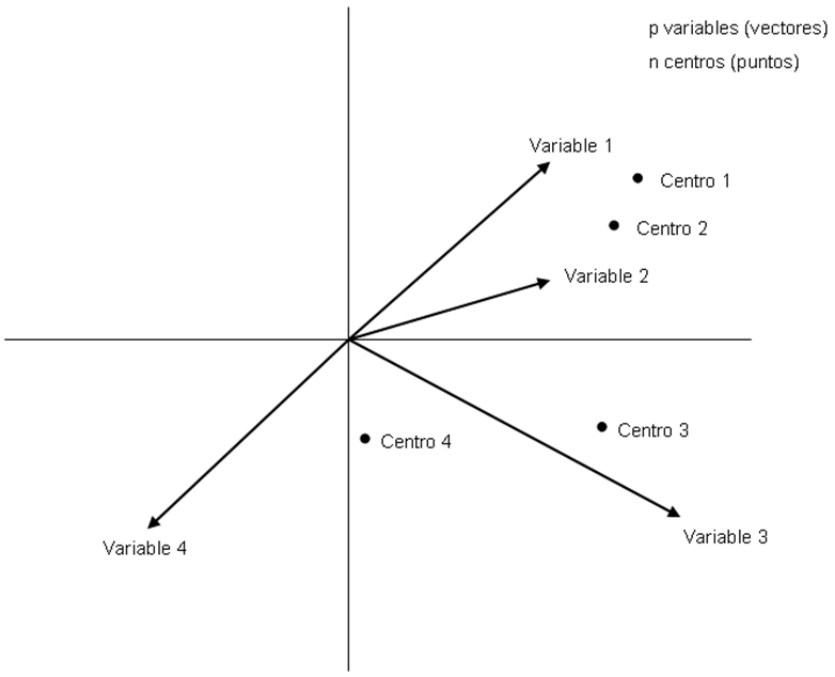

Figura 6.2. Interpretación HJ-Biplot.

Además, como ayuda para una correcta interpretación del gráfico se pueden utilizar algunas medidas adicionales (Galindo y Cuadras, 1986). La correlación al cuadrado entre una variable (indicador) y un factor se interpreta como la contribución relativa del factor al elemento (CRFA). Dado que los factores obtenidos no están correlacionados, la variabilidad de una variable representada en un plano se obtiene sumando las contribuciones de los ejes que conforman el plano, cantidad que se denomina calidad de representación (CLR). Únicamente las variables con CLRs elevadas deben de interpretarse en el plano en cuestión. Una medida análoga de la CLR se puede obtener para cada unidad (centro). Se considera que un centro del CSIC está bien representado cuando se recoge la mayoría de su información (medida a través de la variabilidad) en la dimensión reducida. Debido a que la representación se centra en el origen, la variabilidad de cada centro se mide por su distancia al cuadrado del origen, de modo que la calidad de representación se puede medir por la relación entre la distancia al cuadrado en la dimensión reducida y la distancia al cuadrado en el espacio completo. Geométricamente, es el cuadrado del coseno del ángulo entre el vector en el espacio completo y su proyección sobre el plano de la representación. Si se sumasen las calidades de representación de los ejes factoriales necesarios para absorber toda la variabilidad presente en los datos, la suma de las CLRs sería 1. En este estudio la CLR se valora en una escala de 0 a 1000 puntos.

Conviene resaltar que, a pesar de la aparente similitud entre el HJ-Biplot y el Análisis de Correspondencias, ambos métodos proporcionan resultados diferentes. Ambos métodos pretenden representar las filas y las columnas sobre un subespacio de baja

dimensión en el que sean interpretables sus posiciones relativas, pero las posiciones de los puntos en uno y otro método son diferentes. Las diferencias radican en:

- ✓ La distancia que se define en el hiperespacio: en el Análisis de Correspondencias se trabaja con la distancia Chi-cuadrado y en el HJ-Biplot con la distancia de Mahalanobis. Dos puntos están más cerca, o más lejos, en el hiperespacio, dependiendo de con qué distancia se observen.
- ✓ El peso que se asigna a cada fila/columna: en el Análisis de Correspondencias las líneas de la matriz tienen asignado un peso inversamente proporcional a su total marginal, lo que no ocurre en el HJ-Biplot.

En ambos métodos se trabaja en términos de absorción de inercia, pero inercia es masa por distancia (al cuadrado). Un punto más pesado obviamente viene representado más cerca del origen (si la masa es grande, la distancia tiene que ser corta). Por eso, en un Análisis de Correspondencias las filas (columnas) más frecuentes están cerca del origen y, por tanto, peor representadas. En cambio, en un HJ-Biplot las masas son unitarias, luego inercia es igual a distancia; de forma que a mayor variabilidad, más información, más inercia y más lejos del origen se sitúan los puntos. De esta manera, en las Correspondencias los centros que toman valores más altos, para los diferentes indicadores, son los que vienen peor representados en el gráfico factorial, es decir, serían los centros con menor relevancia en el análisis y sus posiciones en el gráfico factorial podrían ser aparentes. Con el HJ-Biplot sucede justo lo contrario. Los centros propios y mixtos del CSIC en el período 2006-2009 que presentan mayores valores de impacto y colaboración científica son los que tienen más relevancia en el análisis. Obviamente, para nuestro objetivo, el Análisis de Correspondencias no proporciona una solución factorial óptima. También hay que advertir de la gran diferencia existente entre realizar un Análisis de Cluster sobre los datos originales o hacerla, como se hace en este trabajo, sobre las coordenadas del HJ-Biplot. En el primer caso encontraríamos centros con perfiles similares pero no sería posible conocer por qué se han producido esas agrupaciones. En el Análisis de Cluster basado en las coordenadas del Biplot, conocemos además qué indicadores explican las diferentes agrupaciones encontradas.

El análisis se ha llevado a cabo a través del programa MultBiplot desarrollado por Vicente-Villardón (2010) en el entorno de programación orientado a matrices MATLAB. Ejemplos de su aplicación, tanto en su versión integrada en MATLAB como la actual ya compilada, pueden encontrarse en los trabajos de Demey y otros (2008), Vicente-Villardón y otros (2006) o Vicente-Galindo y otros (2011). También existen otras aplicaciones para los métodos Biplot como las desarrolladas en el entorno R por Faria y Demetrio (2011), Nieto-Librero y otros (2011) o Frutos-Bernal y Galindo (2012).

Los datos se han estandarizado por columna debido a las diferentes unidades de medida de las variables. Los centros con CLRs inferiores a 500 puntos no se representan en los gráficos factoriales. Para la selección del tipo de Cluster se han aplicado, con fines exploratorios, Cluster jerárquicos y se ha afinado la solución mediante los métodos no jerárquicos, en concreto, se usó el método K-means y como medida la distancia euclídea. Para la representación factorial se ha tomado el nombre abreviado de los centros y se han clasificado según las ocho áreas científico-técnicas del CSIC (ver Anexo): Humanidades y Ciencias Sociales, Biología y Biomedicina, Recursos Naturales, Ciencias Agrarias, Ciencia y Tecnologías Físicas, Ciencia y Tecnología de Materiales, Ciencia y Tecnología de Alimentos y Ciencia y Tecnologías Químicas (se ha considerado cada centro asignado a su área principal).

6.3. Resultados

La producción científica del CSIC en el periodo 2006-2009 asciende a 28834 artículos. La distribución de la producción por áreas científico-técnicas y el número de centros incluidos en cada área se muestra en la Tabla 6.1.

Tabla 6.1. Número de centros con producción y número de artículos por áreas científico-técnicas del CSIC.

Área CSIC	Nº Centros	Nº Artículos
Humanidades y Ciencias Sociales	19	598
Biología y Biomedicina	24	5345
Recursos Naturales	23	5199
Ciencias Agrarias	14	2448
Ciencia y Tecnologías Físicas	26	5709
Ciencia y Tecnología de Materiales	11	5252
Ciencia y Tecnología de Alimentos	8	1804
Ciencia y Tecnologías Químicas	15	3743
Total	140	28834

* Nota: el sumatorio de artículos es superior al total real porque existe colaboración entre centros de distintas áreas.

Se han retenido tres ejes pues se consigue una inercia acumulada muy elevada, 91,1%, más que suficiente para caracterizar, con garantías, la actividad científica de los centros propios y mixtos del CSIC en relación a las variables de impacto y colaboración consideradas (ver Tabla 6.2).

Tabla 6.2. Valores propios y varianza explicada.

Ejes	Inercia		
	Valor propio	**Var. Explicada**	**Var. Acumulada**
1	20,01	49,43	49,43
2	13,74	23,31	72,74
3	12,20	18,37	91,11
4	7,34	6,66	97,77
5	4,26	2,24	100

Atendiendo a las contribuciones del factor al elemento para las columnas (ver Tabla 6.3), se observa que todas las variables, han de interpretarse en el plano factorial 1-2, a excepción del porcentaje de artículos sin colaboración, que queda mejor recogido en el plano 1-3. La PN, aunque presenta contribuciones ligeramente superiores en el plano 1-3, se analiza en el plano 1-2 por resultar de mayor interés su interpretación junto al resto de indicadores recogidos en dicho plano. En cuanto a las filas, de los 136 centros del CSIC tan sólo seis no han quedado bien recogidos en los tres primeros ejes: Instituto de Ciencia y Tecnología de Polímeros, Instituto de Recursos Naturales y Agrobiología de Sevilla, Instituto de Microbiología Bioquímica, Instituto de Física Fundamental, Instituto de Ciencia y Tecnología de Alimentos y Nutrición y Centro de Investigaciones Físicas Isla de Cartuja[17].

Tabla 6.3. Calidad de representación paras las columnas.

Variables	Eje 1	Eje 2	Eje 3
PN Media	**582**	81	213
% Art. Q1	**775**	85	42
Citas Relativas Mundo	**665**	0	45
% Art. sin colaboración	**446**	15	**535**
% Art. col. nacional	17	**852**	124
% Art. col. internacional	**481**	**366**	143

6.3.1. Análisis del impacto y la colaboración: plano 1-2

En la Figura 6.3 se muestra el gráfico factorial del plano 1-2, donde la inercia acumulada asciende al 72,7%. Los indicadores bibliométricos analizados se representan mediante vectores, mientras que los centros se identifican mediante puntos, cuya etiqueta incluye el nombre abreviado del centro (ver Anexo) y su color varía en función de su área de pertenencia. Los vectores que no están bien

[17] Nótese que Isla de Cartuja es un centro que incluye varios institutos, también visibles en este trabajo. Sólo se asignan al centro aquellos artículos firmados por el centro que no incluyen ninguno de sus institutos.

representados en este plano se presentan de un color más claro (porcentaje de artículos sin colaboración).

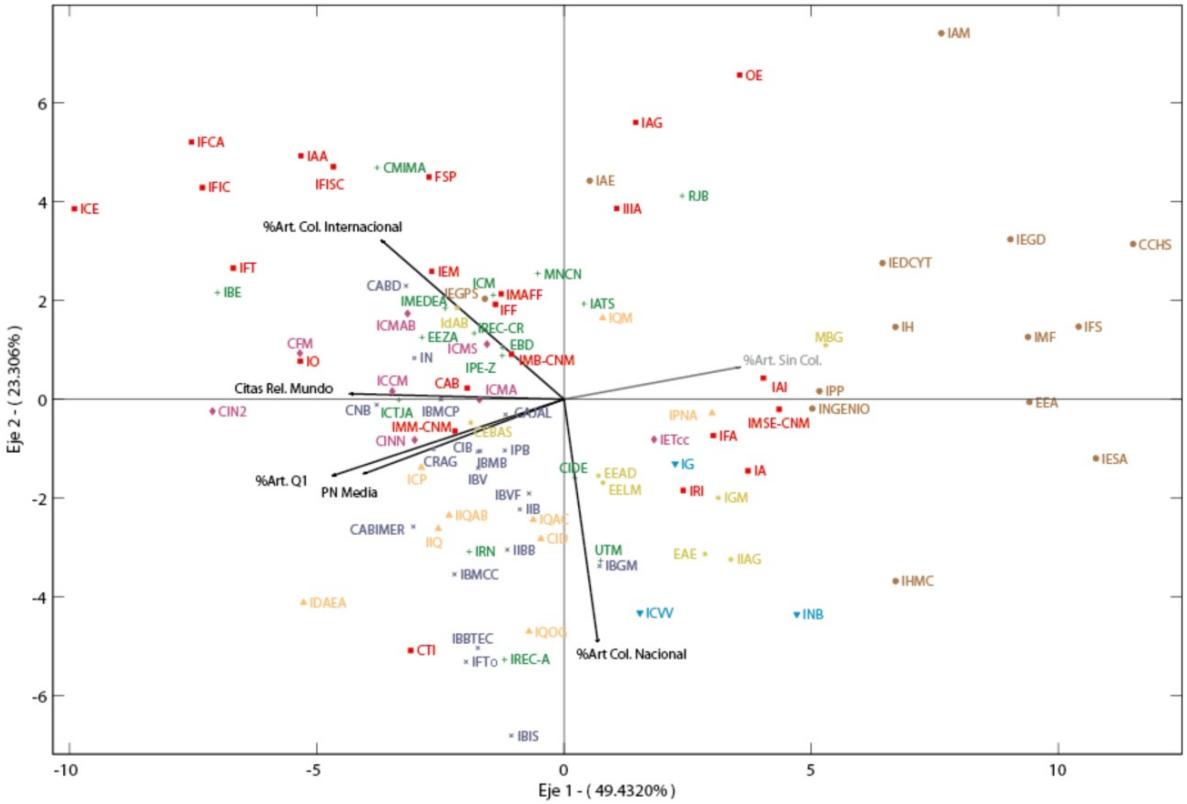

Figura 6.3. Representación factorial resultante del HJ-Biplot, plano 1-2.

Notas: Orden de los cuadrantes: 1º - superior derecho, 2º - superior izquierdo, 3º - inferior izquierdo, 4º - inferior derecho. Se representan los centros con calidades de representación ≥ 500 puntos.

Leyenda con colores de áreas: Humanidades y Ciencias Sociales, Biología y Biomedicina, Recursos Naturales, Agrarias, Físicas, Materiales, Alimentos, Química.

Los indicadores bibliométricos, representados mediante vectores, indican que existe una relación fuerte y directa entre el porcentaje de artículos Q1 en JCR, la PN de la revista en la que se ha publicado y las CRM, siendo esta última variable la más importante para discriminar entre los centros en el eje 1. Igualmente, hay cierta correlación entre el porcentaje de artículos en colaboración internacional y las CRM (ángulo agudo entre los vectores). También cabe señalar que el porcentaje de artículos en colaboración nacional es la variable más importante para discriminar entre individuos en el eje 2, siendo además el indicador que ha presentado una mayor variabilidad en el estudio. Asimismo, esta variable correlaciona inversamente con la producción en colaboración internacional (ángulo casi llano entre ambas variables). Por tanto, los centros situados en la parte izquierda del gráfico factorial publican sus resultados de investigación en revistas de primer nivel y gozan de una visibilidad elevada. Además, se pueden diferenciar en función del alcance de la colaboración: para los centros del segundo cuadrante predomina la asociación con autores adscritos

a centros extranjeros, mientras que la pauta más común para los centros del tercer cuadrante es la coautoría nacional. La variable restante, porcentaje de artículos sin colaboración, y varios centros del área de Alimentos presentan mayores contribuciones en el plano 1-3.

En algunos casos, los centros de una misma área tienden a situarse en la misma zona del gráfico HJ Biplot, lo que indica que presentan características similares en lo que respecta a colaboración e impacto. Es el caso de los centros de Humanidades y Ciencias Sociales (primer cuadrante), Biología y Biomedicina y Químicas (tercer cuadrante) o Alimentos (cuarto cuadrante). Sin embargo, otras áreas son más heterogéneas en lo que respecta al comportamiento de sus centros. Es el caso de Físicas, Ciencias Agrarias o Recursos Naturales, en las que los centros presentan una mayor dispersión en su comportamiento.

En términos generales, atendiendo a la posición de los centros y las variables en el gráfico factorial, se observa que las áreas CSIC que tienden a mostrar un mayor impacto de su producción son Físicas, Químicas y, en menor medida, Biología y Biomedicina, Recursos Naturales y Materiales. Destacan por su alto porcentaje de artículos en revistas Q1 el Instituto de Física Teórica (IFT) (86,5%), el Instituto de Diagnóstico Ambiental y Estudios del Agua (IDAEA) (85,6%) y el Instituto de Biología Evolutiva (IBE) (82,2%). Si se toma como indicador las CRM destacan algunos centros del área de Física como el Instituto de Física Corpuscular (IFIC) o el Instituto de Física de Cantabria (IFCA), junto a algún centro de otras áreas como el Centre d´Investigació en Nanociència i Nanotecnologia (CIN2) del área de Materiales. La PN no discrimina bien entre los centros con valores altos en sus indicadores de impacto y colaboración (el 50% presenta PN comprendidas entre 0,66-0,77), pero es un indicador útil para caracterizar a los centros que han publicado una parte importante de su producción en revistas de bajo factor de impacto dentro de sus respectivas disciplinas, como el Centro de Ciencias Humanas y Sociales[18] (CCHS) (PN=0,41) o el Instituto de Economía, Geografía y Demografía (IEGD) (PN=0,43) adscritos a Humanidades y Ciencias Sociales. Esto es más habitual en centros que publican principalmente en revistas españolas, tal como sucede en Humanidades y Ciencias Sociales, por la peor posición que suelen ocupan estas revistas atendiendo al factor de impacto.

En cuanto a la colaboración, los centros situados en el segundo cuadrante se caracterizan por un elevado número de artículos en colaboración internacional. Entre ellos se puede nombrar el Instituto de Física de Cantabria (IFCA), el Instituto de Astrofísica de Andalucía (IAA), el Centro Mediterráneo de Investigaciones Marinas y Ambientales (CMIMA), el Instituto de Ciencias del Espacio (ICE) y el Instituto de Física Corpuscular (IFIC) adscritos a Físicas (salvo el CMIMA que pertenece al área de

[18] Incluye siete institutos que se analizan de forma independiente en este estudio. Sólo se asignan al CCHS los artículos firmados únicamente por el centro sin incluir ninguno de sus institutos.

Recursos Naturales), que han publicado más del 80% de su producción en coautoría con autores adscritos a centros extranjeros. Resulta llamativa la presencia en esta zona del gráfico de dos centros de Humanidades y Ciencias Sociales, el Instituto de Análisis Económico (IAE) y el Instituto de Estudios Gallegos Padre Sarmiento (IEGPS), que presentan una alta colaboración internacional, 66,7% y 63,6% respectivamente, poco habitual en el área, ya que los centros restantes se sitúan en la parte derecha del gráfico con altas tasas de documentos sin colaboración. Otras áreas como Ciencias Agrarias o Químicas se caracterizan por una mayor investigación en colaboración nacional. En el caso de Biología y Biomedicina coexisten centros con una alta actividad en colaboración nacional, como el Instituto de Biomedicina de Sevilla (IBIS), tercer cuadrante de la Figura 6.3, que ha publicado el 65% de sus artículos en colaboración nacional; y centros con alta actividad en colaboración internacional, como el Centro Andaluz de Biología del Desarrollo (CABD) o el Instituto de Neurociencias de Alicante (IN), segundo cuadrante, que han publicado en coautoría internacional más del 50% de sus documentos. Estos centros se sitúan en la parte izquierda del gráfico factorial porque el impacto de su producción tiende a ser alto. Por el contrario, el Instituto de Nutrición y Bromatología (INB), de Alimentos, y el Instituto de Historia de la Medicina y de la Ciencia López Piñero (IHMC), de Humanidades y Ciencias Sociales, se sitúan en la parte inferior del cuarto cuadrante pues publican, principalmente, en colaboración nacional pero su impacto en la comunidad científica es menor.

6.3.2. Análisis del impacto y la colaboración: plano 1-3

La absorción de inercia en el plano factorial 1-3 es del 67,8% (ver Figura 6.4). Esta representación resulta de interés para caracterizar el porcentaje de artículos sin colaboración, al ser óptima la calidad de representación para este indicador.

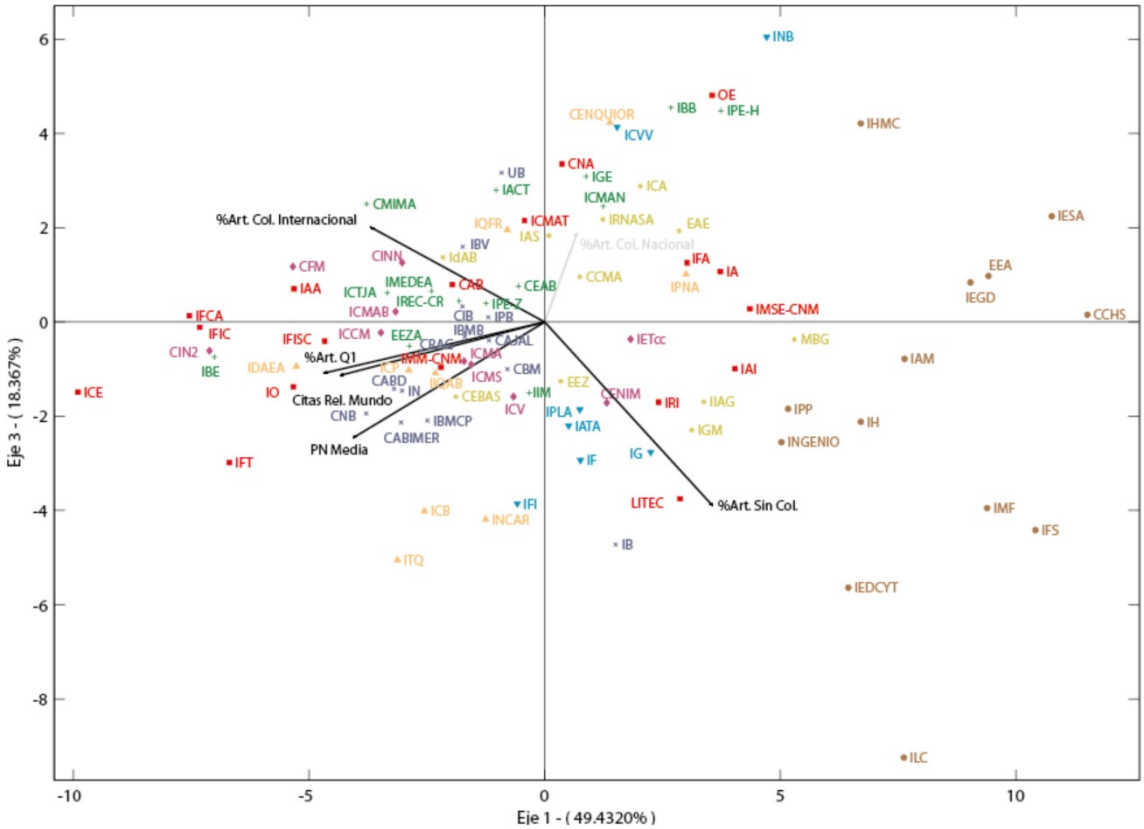

Figura 6.4. Representación factorial resultante del HJ-Biplot, plano 1-3.

Notas: Orden de los cuadrantes: 1º - superior derecho, 2º - superior izquierdo, 3º - inferior izquierdo, 4º - inferior derecho. Se representan los centros con calidades de representación ≥ 500 puntos.

Leyenda con colores de áreas: Humanidades y Ciencias Sociales, Biología y Biomedicina, Recursos Naturales, Agrarias, Físicas, Materiales, Alimentos, Química.

Se observa una relación inversa entre el porcentaje de artículos sin colaboración y los realizados en coautoría internacional. Esta variable parece independiente de los indicadores relativos a visibilidad e impacto. Los centros situados en el cuarto cuadrante presentan un patrón de publicación en el que predominan los artículos firmados por un único centro, práctica común entre los centros del área de Humanidades y Ciencias Sociales (Gómez y otros, 2011b). Destacan por su situación algo periférica en el cuarto cuadrante algunos institutos con más de 2/3 partes de su producción firmada por un solo centro, como son el Instituto de Lenguas y Culturas del Mediterráneo y Oriente Próximo (ILC) (87%), el Instituto de Filosofía (IFS) (73,7%), la Institución Milá Fontanals (IMF) (68,4%) y el Instituto de Estudios Documentales sobre Ciencia y Tecnología (IEDCYT) (72,9%). En los institutos más próximos al centro de gravedad comienza a incrementarse la actividad en colaboración y los resultados de investigación se publican en revistas de mayor visibilidad internacional, así el Instituto de Gestión de la Innovación y del Conocimiento (INGENIO), el Instituto de Políticas y Bienes Públicos (IPP) han publicado un 44,2% y un 38,5% de artículos en revistas indexadas en el Q1. Este plano recoge, además, a varios centros del área de Alimentos

que no quedaban bien representados en el plano factorial 1-2. Se observa que la mayor parte de los institutos de esta área muestran alta actividad sin colaboración, situándose en el cuarto cuadrante de la Figura 6.4; a excepción de dos institutos con escasa actividad sin colaboración, pero con alta colaboración nacional, que se sitúan en el primer cuadrante. Finalmente, es interesante señalar la presencia de algunos centros como la Misión Biológica de Galicia (MBG), de Agrarias, o el Instituto de Automática Industrial (IAI), de Físicas, que presentan una alta actividad sin colaboración que los diferencia de otros centros de sus áreas respectivas.

6.3.3. Clusters según tipo de colaboración

A través de las coordenadas Biplot se han calculado los Clusters (método K-means, distancia euclídea). Se observa en el gráfico factorial (Figura 6.5) que los centros forman conglomerados en función de su comportamiento en las variables de impacto y colaboración. Las calidades de representación para cada conglomerado en el plano 1-2 se exponen en la Tabla 6.4.

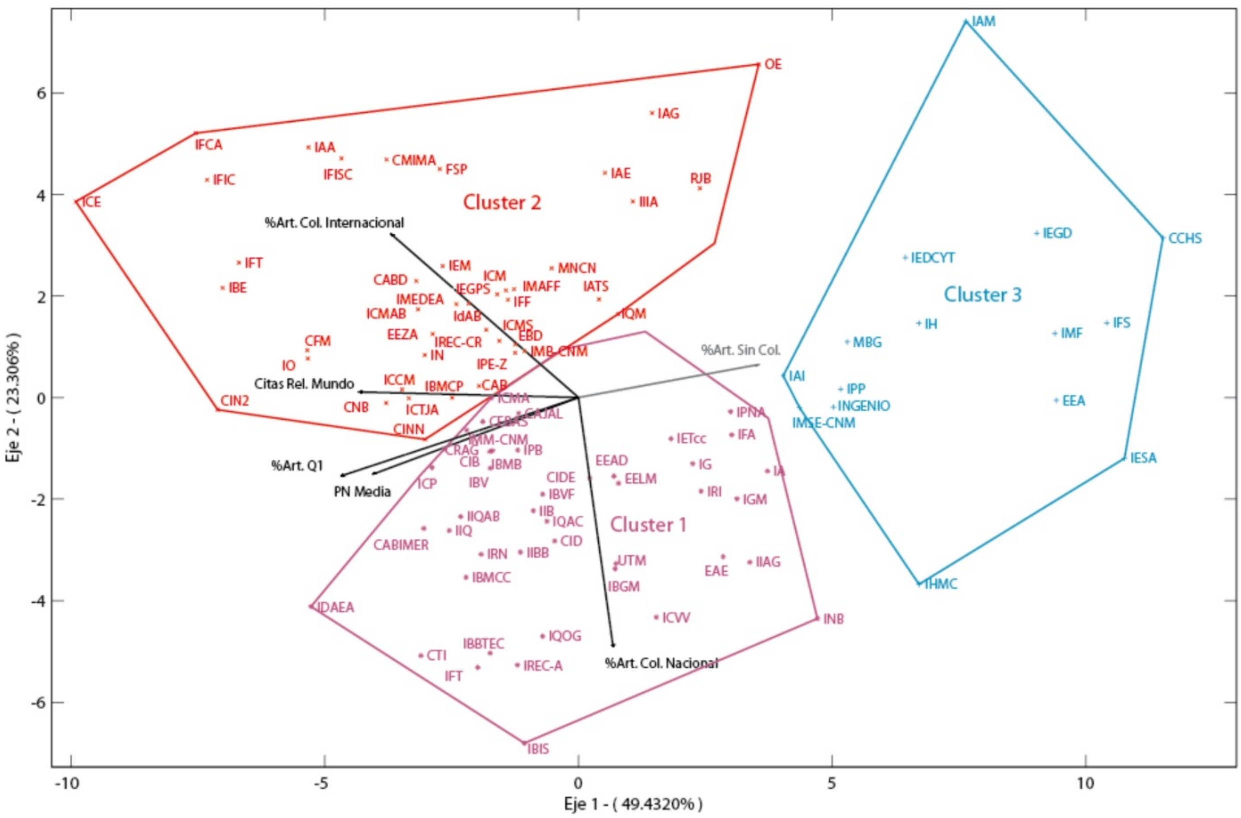

Figura 6.5. Representación factorial del HJ-Biplot por Clusters, plano 1-2.

Notas: se representan los centros con calidades de representación ≥ 500 puntos.

Tabla 6.4. Conglomerados y calidades de representación en el plano 1-2

	n	CLR - 1	CLR - 2	CLR 1-2
Cluster 1	73	0,23	99,24	99,47
Cluster 2	47	53,77	43,82	97,59
Cluster 3	16	93,41	2,48	95,89

La descripción general de los clusters en función de las seis variables utilizadas se muestra en la Tabla 6.5. Observamos importantes diferencias entre clusters en las prácticas de colaboración predominantes en cada caso, de forma que en los centros del Cluster 1 predomina la colaboración de orientación nacional, mientras que la colaboración internacional predomina en el Cluster 2, y la producción sin colaboración entre centros impera en el Cluster 3. La Figura 6.6 muestra la composición de los clusters atendiendo al área científico-técnica de sus centros. Considerando la distribución total de centros por áreas, observamos que el Cluster 1 se caracteriza por una alta presencia relativa de los centros de las áreas de Biología y Biomedicina, Ciencias Agrarias, Alimentos y Químicas; y una ausencia de centros de Humanidades y Ciencia Sociales. Así, en el Cluster 1 predominan los centros de Biología y Biomedicina (26%), Químicas (19%) y Agrarias (16%). Todos los centros e institutos del área de Alimentos quedan integrados en el Cluster 1. En el Cluster 2 predominan los centros adscritos a Físicas (36%), Recursos Naturales (30%) y Materiales (15%), que muestran una alta presencia relativa. En el Cluster 3 predominan las Ciencias Sociales y Humanidades (81%).

Tabla 6.5. Descriptivos de los indicadores de impacto y colaboración según Clusters.

	PN	Q1	CRM	% sin colaboración	% colaboración nacional	% colaboración internacional
Cluster 1	0,73 ± 0,01	59,19 ± 1,44	1,20 ± 0,04	19,97 ± 1,44	43,62 ± 1,31	36,41 ± 1,20
Cluster 2	0,73 ± 0,01	59,09 ± 2,00	1,43 ± 0,08	13,17 ± 1,05	26,02 ± 1,40	60,81 ± 1,78
Cluster 3	0,60 ± 0,03	20,15 ± 3,49	0,60 ± 0,09	51,20 ± 4,37	30,95 ± 4,11	17,85 ± 2,96
Total	0,72 ± 0,01	54,56 ± 1,55	1,21 ± 0,04	21,29 ± 1,39	36,05 ± 1,21	42,66 ± 1,56

Nota: datos expresados como media ± desviación típica.

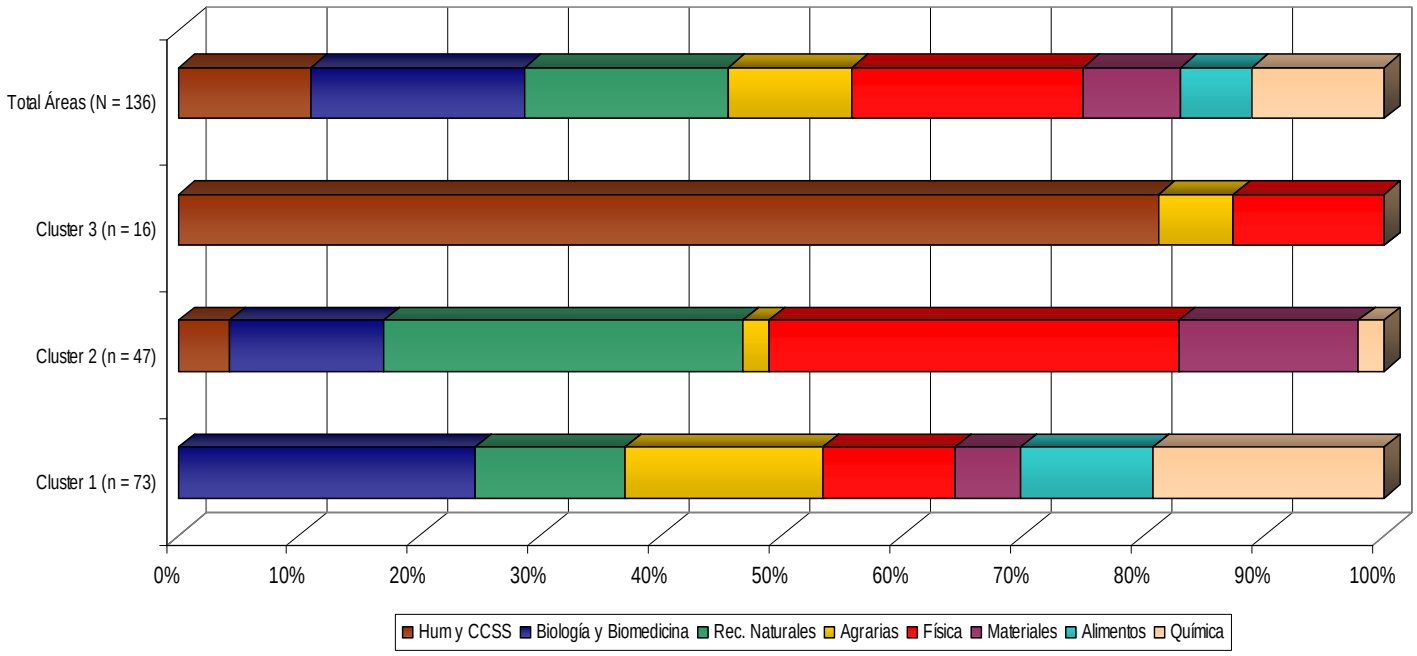

Figura 6.6. Configuración de los conglomerados por áreas.

En promedio, los Clusters 1 y 2 presentan mayores valores de impacto que el Cluster 3. En la Figura 6.5 se identifican los centros con más impacto en la parte izquierda de los Clusters 1 y 2. En el Cluster 3 los centros tienen en promedio menos impacto y se caracterizan por la publicación de artículos firmados por un único centro así como por englobar la mayor parte de los centros de Humanidades y Ciencias Sociales (todos excepto el IAE y el IEGPS antes mencionados). Resulta llamativa la presencia en el Cluster 3 del Instituto de Microelectrónica de Sevilla (IMSE-CNM) y el Instituto de Automática Industrial (IAI), del área de Físicas, así como de la Misión Biológica de Galicia (MBG), del área de Ciencias Agrarias, que se caracterizan por su alta actividad sin colaboración, baja colaboración internacional, y bajo porcentaje de documentos en revistas Q1 -comparado con la media de sus áreas-.

6.4. Discusión y conclusiones

En este estudio se ha analizado la distribución de los institutos y centros del CSIC en función de sus prácticas de colaboración y su impacto mediante la técnica HJ-Biplot que permite realizar una representación gráfica de datos multivariantes en el que filas y columnas pueden ser superpuestas en un mismo sistema de referencia con máxima calidad de representación. Se pone de manifiesto la relación entre variables, observándose una correlación fuerte y directa entre los indicadores basados en citas y factor de impacto, de forma que el número de citas recibidas va asociado a un mayor alcance de la colaboración. Así, el estudio permite observar que las características analizadas son, en parte, dependientes de cada área, ya que los centros e institutos de

cada área tienden a posicionarse en la misma zona de los gráficos factoriales resultantes. Sin embargo, existe también cierta heterogeneidad intra-área, de forma que Humanidades y Ciencias Sociales, y Alimentos muestran la mayor homogeneidad, mientras que Físicas y Agrarias presentan mayor variedad de comportamientos entre sus centros e institutos.

Los resultados obtenidos son consistentes con otros estudios que señalan la influencia positiva de la colaboración internacional sobre el impacto de la investigación (ver por ejemplo, Glänzel y Schubert, 2001). El análisis de los datos ha puesto de manifiesto que las áreas CSIC de un mayor impacto y visibilidad internacional son Físicas, caracterizada por la coautoría internacional; y Químicas, donde la producción con autores adscritos a centros nacionales es el patrón más común. Las mayores tasas de colaboración internacional se dan en las áreas de Físicas y Recursos Naturales. En el caso de Ciencias Sociales y Humanidades, se aprecian diferencias entre los centros de Humanidades, por un lado, con un escaso ratio de colaboración e impacto; y los centros más próximos a Ciencias Sociales, por otro lado, que tienden a presentar mayor colaboración y visibilidad internacional. No obstante, hay que señalar que los resultados aquí expuestos sobre Humanidades sólo representan un pequeño fragmento de la actividad científica del área, pues sus centros publican una parte importante de sus resultados en libros y revistas de ámbito regional no siempre incluidas en WoS y, por otro lado, la base de datos no calcula el factor de impacto para las revistas de Humanidades (González-Albo y otros, 2012).

El estudio actual permite analizar el comportamiento específico de cada centro, y su situación en el contexto de su área y del total de los centros e institutos del CSIC. Así, es posible identificar centros que se salen del patrón general de su área, y determinar qué faceta de su actividad les hace singulares. Exponemos a continuación algunos ejemplos en varias de las áreas analizadas.

En Humanidades y Ciencias Sociales hay que destacar el Instituto de Análisis Económico (IAE) y el Instituto de Estudios Gallegos Padre Sarmiento (IEGPS) que se sitúan en el segundo cuadrante de la Figura 6.3, lejos del resto de institutos de su área ubicados principalmente en el primer cuadrante; lo que se explica por su alta orientación internacional. El IAE participa en diversos proyectos europeos (CSIC, 2009) y publica prácticamente toda su producción en revistas internacionales, mientras que el IEGPS tiene una orientación más regional y su producción en WoS sólo representa 1/3 de su actividad científica (Gómez y otros, 2011a), pero es la que queda reflejada en el estudio actual. También en el área de Humanidades y Ciencias Sociales llama la atención por su distante posición en el cuarto cuadrante el Instituto de Historia de la Medicina y de la Ciencia López Piñero (IHMC), cuya actividad se caracteriza por una alta orientación a la colaboración nacional (68%), pero situando su producción en revistas de poca visibilidad (25% en el primer cuartil).

Los institutos de Recursos Naturales tienden a concentrarse en el segundo cuadrante de la Figura 6.3. No obstante, algunos institutos con un comportamiento "atípico" o singular se sitúan en otras zonas del gráfico. Cabe señalar al Real Jardín Botánico (RJB), situado en el primer cuadrante, con una tasa reseñable de colaboración internacional (59%) pero con poca producción en revistas del primer cuartil (33%). Un análisis detallado de los datos pone en evidencia que sus publicaciones tienden a concentrarse en revistas muy especializadas, sobre todo de botánica, alejadas de las primeras posiciones ocupadas por revistas de ámbito más general. Por otro lado, en el tercer y cuarto cuadrante se sitúan algunos institutos que desarrollan una investigación con un carácter más aplicado, como son el Instituto de Investigación en Recursos Cinegéticos de Albacete (IREC-A) que presenta una alta tasa de colaboración nacional (69%) y una visibilidad notable (65% de artículos en el primer cuartil y una PN de 0,71), y la Unidad de Tecnología Marina (UTM), que presta un servicio de apoyo logístico y técnico (CSIC, 2009), suele colaborar con centros nacionales (61%) y tiene un visibilidad algo menor (47% de artículos en primer cuartil y PN=0,68).

Aunque los institutos de Materiales tienden a concentrarse en el segundo cuadrante de la Figura 6.3, destaca la situación en el cuarto cuadrante del Instituto de Ciencias de la Construcción Eduardo Torroja (IETcc). Este centro se aleja del patrón de su área debido a un menor ratio de colaboración internacional (32%), lo que puede estar asociado al desarrollo de una investigación más aplicada, ya que presta servicios de apoyo científico-técnico al sector de la construcción (CSIC, 2009).

Los institutos de Ciencias Agrarias aparecen bastante dispersos, aunque predominan en la parte derecha de la Figura 6.3. Sin embargo, en la mitad izquierda se identifican dos institutos que sobresalen por su alta actividad en revistas Q1 y colaboración internacional, como son el Instituto de Agrobiotecnología de Navarra (IdAB), ubicado en el segundo cuadrante, y el Centro de Edafología y Biología Aplicada del Segura (CEBAS), en el tercer cuadrante (presenta menor colaboración internacional que el IdAB).

El área de Físicas incluye institutos dispersos por los cuatro cuadrantes de la Figura 6.3. En el segundo cuadrante se incluyen centros orientados a la "Big Science", como el Instituto de Ciencias del Espacio (ICE) o el Instituto de Astrofísica de Andalucía (IAA), que tienen una alta orientación internacional de la investigación, que se publica en revistas de alto factor de impacto. Por el contrario, en el cuarto cuadrante se sitúan centros que trabajan en la línea de tecnologías físicas o informáticas (CSIC, 2009) y cuya investigación tiene una orientación más nacional, como el Instituto de Acústica (IA).

En conclusión, el HJ-Biplot se ha revelado como una herramienta multivariante sumamente útil en el análisis de datos bibliométricos en la etapa descriptiva. Este método, frente al Análisis de Correspondencias -una potente técnica de ordenación

pensada para tablas de contingencia aunque aplicable a matrices de datos en las que tenga sentido trabajar con perfiles-, presenta como principales ventajas un ámbito de aplicación mucho más general al ser aplicable a cualquier matriz de datos y la posibilidad de detectar qué indicadores bibliométricos son los responsables de las agrupaciones de los centros. La aplicación del HJ-Biplot al estudio de la producción científica del CSIC nos ha permitido caracterizar la actividad de las áreas en cuanto a colaboración e impacto se refieren e identificar centros con un comportamiento sobresaliente o singular que los diferencia del resto de su área.

Agradecimientos

Agradecemos los comentarios de Isabel Gómez Caridad sobre una versión previa de este documento. Adrián A. Díaz-Faes cuenta con una beca predoctoral de la Junta de Ampliación de Estudios – Consejo Superior de Investigaciones Científicas (JAE-CSIC). Este artículo ha sido realizado en el marco de los proyectos 200410E605 y CSO2008-06310.

Referencias

Abd el Kader, M., Ojasoo, T., Miquel, J.F., Okubo, Y., & Doré, J.C. (1998). Hierarchical author networks: an analysis of European Molecular Biology Laboratory (EMBL) publications. *Scientometrics*, *42*(3), 405-421.

Anuradha, K.T., & Urs, S.R. (2007). Bibliometric indicators of Indian research collaboration patterns: a correspondence analysis. *Scientometrics*, *71*(2), 179-189.

Benzécri, J.P. (1973). *L'analyse de Données. 2. L'analyse des correspondances*. Paris: Dunod.

Bordons, M., & Barrigón, S. (1992). Bibliometric analysis of publications of Spanish pharmacologists in the SCI (1984-1989) Part I. *Scientometrics*, *25*(3), 425-446.

Bordons, M., Bravo, C., & Barrigón, S. (2004). Time-tracking of the research profile of a drug using bibliometric tools. *Journal of the American Society for Information Science and Technology*, *55*(5), 445-461.

Caballero-Juliá, D. (2011). *El HJ-Biplot como herramienta en el análisis de grupos de discusión. Calidad de vida en la ludopatía: una propuesta sociológica.* (Tesis de maestría). http://hdl.handle.net/10366/108778.

Cárdenas, O., Galindo, M.P., & Vicente-Villardón, J.L. (2007). Los métodos Biplot: evolución y aplicaciones. *Revista Venezolana de Análisis de Coyuntura*, *13*(1), 279-303.

Díaz-Faes, A.A., Benito-García, N., Martín-Rodero, H., & Vicente-Villardón, J.L. (2011). Propuesta de aplicabilidad del método multivariante gráfico Biplot a los estudios bibliométricos en biomedicina. *Actas XIV Jornadas Nacionales de Información y*

Documentación en Ciencias de la Salud, p. 66. Cádiz, España: Biblioteca Virtual del Sistema Sanitario Público de Andalucía.

Doré, J.C., Dutheuil, C., & Miquel, J.F. (2000). Multidimensional analysis of trends in patent activity. *Scientometrics*, *47*(3), 475-492.

Doré, J.C., & Ojasoo, T. (2001). How to analyze publication time trends by correspondence factor analysis: Analysis of publications by 48 countries in 18 disciplines over 12 years. *Journal of the American Society for Information Science and Technology*, *52*(9), 763-769.

Dore, J.C., Ojasoo, T., Okubo, Y., Durand, T., Dudognon, G., & Miquel, J.F. (1996). Correspondence factor analysis of the publication patterns of 48 countries over the period 1981-1992. *Journal of the American Society for Information Science*, *47*(8), 588-602.

Faria, J.C., & Demetrio, C.G.B. (2011). BPCA: Biplot of multivariate data based on Principal Components Analysis [Programa informático]. ESALQ, USP, Brasil. http://cran.r-project.org/web/packages/bpca/citation.html

Frutos-Bernal, E. (2012). GGEBiplotGUI: Interactive GGE Biplots in R [Programa informático]. Salamanca, España: Departamento de Estadística, Universidad de Salamanca. http://cran.r-project.org/web/packages /GGEBiplotGUI/index.html

Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*(3), 453-467.

Gabriel, K.R., & Odoroff, C.L. (1990). Biplots in biomedical research. *Statistics in Medicine*, *9*, 469-485.

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

Galindo, M.P., & Cuadras, C. (1986). *Una extensión del método Biplot y su relación con otras técnicas*. Publicaciones de Bioestadística y Biomatemática, 17. Barcelona: Universidad de Barcelona.

Glanzel, W., & Schubert, A. (2001). Double effort = double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, *50*(2), 199-214.

Gómez, I., Bordons, M., Morillo, F., Moreno, L., Aparicio, J., Díaz-Faes, A.A., & González-Albo, B. (2011a). *La actividad científica del CSIC a través de indicadores bibliométricos (Web of Science, 2006-2010)*. Madrid: IEDCYT, CCHS, CSIC. http://hdl.handle.net/10261/48118.

Gómez, I., Bordons, M., Morillo, F., Moreno, L., & González-Albo, B. (2011b). *La actividad científica del CSIC: Indicadores de producción e impacto por tipo de colaboración (WoS, 2004-2009)*. Madrid: IEDCYT, CCHS, CSIC. http://hdl.handle.net/ 10261/38113.

González-Albo, B., Moreno, L., Morillo, F., & Bordons, M. (2012). Indicadores bibliométricos para el análisis de la actividad de una institución multidisciplinar: el CSIC. *Revista Española de Documentación Científica*, *35*(1), 9-37.

Miquel, J.F., Ojasoo, T., Okubo, Y., Paul, A., & Doré, J.C. (1995). World science in 18 disciplinary areas: comparative evaluation of the publication patterns of 48 countries over the period 1981–1992. *Scientometrics*, *33*(2), 149–167.

Moed, H.F. (2005). *Citation analysis in research evaluation*. Dordrecht, The Netherlands: Springer.

Morillo, F., Aparicio, J., González-Albo, B., & Moreno, L. (2012). Towards the automation of address identification. *Scientometrics, 94*(1), 207-224. doi: 10.1007/s11192-012-0733

Nagpaul, P.S. (1995). Contribution of Indian Universities to the mainstream scientific literature: a bibliometric assessment. *Scientometrics*, *32*(1), 11-36.

Nieto-Librero, A.N., Baccala, N., & Galindo, M.P. (2011) MultibiplotGUI: Multibiplot Analysis in R [Programa informático]. Salamanca, España: Departamento de Estadística, Universidad de Salamanca. http://cran.r-project.org/web/packages/multibiplotGUI/index.html

Okubo, Y., Doré, J.C., Ojasoo, T., & Miquel, J.F. (1998). A multivariate analysis of publication trends in the 1980s with special reference to South-East Asia. *Scientometrics*, *41*(3), 273-289.

Pritchard, A. (1969) Statistical bibiography or bibliometrics? *Journal of Documentation*, *25*, 348-349.

Sanz-Casado, E., & Conforti, N. (2005). Análisis de la actividad científica de la Facultad de Humanidades de la Universidad de Mar de Plata, durante el periodo 1998-2001. *Revista Española de Documentación Científica*, *28*(2), 196-205.

Suárez-Balseiro, C., García-Zorita, C., & Sanz-Casado, E. (2009). Multi-authorship and its impact on the visibility of research from Puerto Rico. *Information Processing and Management*, *45*, 469-476.

Subramanyam, K. (1983). Bibliometric studies of research collaboration: a review. *Journal of Information Science*, *6*, 33-38.

Vicente-Tavera, S. (1992). *Las técnicas de representación de datos multidimensionales en el estudio del índice de producción industrial en la C.E.E.* (Tesis doctoral no publicada). Universidad de Salamanca, Departamento de Estadística, Salamanca España.

Vicente-Villardón, J.L. (2010). Multbiplot: a packaged for multivariate analysis using Biplots. (versión 1.0) [Programa informático]. Salamanca, España: Departamento de Estadística, Universidad de Salamanca. http://biplot.dep.usal.es/classicalbiplot/.

Anexo. Relación de centros propios y mixtos del CSIC por áreas científico-técnicas (2006-2009)

Centros
Área 1. Humanidades y Ciencias Sociales
Centro de Ciencias Humanas y Sociales (CCHS), Madrid
Escuela de Estudios Árabes (EEA), Granada
Escuela de Estudios Hispanoamericanos (EEHA), Sevilla
Escuela Española de Historia y Arqueología de Roma (EEHAR)
Institución Milá y Fontanals (IMF), Barcelona
Instituto de Análisis Económico (IAE), Barcelona
Instituto de Arqueología (IAM), Mérida
Instituto de Economía, Geografía y Demografía (IEGD), Madrid
Instituto de Estudios Documentales sobre la Ciencia y la Tecnología (IEDCYT), Madrid
Instituto de Estudios Gallegos "Padre Sarmiento" (IEGPS), A Coruña
Instituto de Estudios Islámicos y del Oriente Próximo (IEIOP), Zaragoza
Instituto de Estudios Sociales Avanzados de Andalucía (IESA), Córdoba
Instituto de Filosofía (IFS), Madrid
Instituto de Gestión de la Innovación y del Conocimiento (INGENIO), Valencia
Instituto de Historia (IH), Madrid
Instituto de Historia de la Medicina y de la Ciencia "López Piñero" (IHMC), Valencia
Instituto de Lengua, Literatura y Antropología (ILLA), Madrid
Instituto de Lenguas y Culturas del Mediterráneo y Oriente Próximo (ILC), Madrid
Instituto de Políticas y Bienes Públicos (IPP), Madrid
Área 2. Biología y Biomedicina
Centro Andaluz de Biología del Desarrollo (CABD), Sevilla
Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER), Sevilla
Centro de Biología Molecular "Severo Ochoa" (CBM), Madrid
Centro de Investigación Cardiovascular (CIC), Barcelona
Centro de Investigación en Agrogenómica (CRAG), Barcelona
Centro de Investigaciones Biológicas (CIB), Madrid
Centro Nacional de Biotecnología (CNB), Madrid
Instituto de Biología Molecular de Barcelona (IBMB)
Instituto de Biología Molecular y Celular de Plantas (IBMCP), Valencia
Instituto de Biología Molecular y Celular del Cáncer (IBMCC), Salamanca
Instituto de Biología y Genética Molecular (IBGM), Valladolid
Instituto de Biomedicina de Sevilla (IBIS)
Instituto de Biomedicina de Valencia (IBV)
Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Santander
Instituto de Bioquímica (IB), Madrid
Instituto de Bioquímica Vegetal y Fotosíntesis (IBVF), Sevilla
Instituto de Farmacología y Toxicología (IFTo), Madrid
Instituto de Investigaciones Biomédicas "Alberto Sols" (IIB), Madrid
Instituto de Investigaciones Biomédicas de Barcelona (IIBB)
Instituto de Microbiología Bioquímica (IMB), Salamanca
Instituto de Neurobiología "Ramón y Cajal" (CAJAL), Madrid
Instituto de Neurociencias (IN), Alicante
Instituto de Parasitología y Biomedicina “López-Neyra” (IPB), Granada
Unidad de Biofísica (UB), Leioa [Vizcaya]
Área 3. Recursos Naturales
Centro de Estudios Avanzados de Blanes (CEAB), Girona
Centro de Investigaciones sobre Desertificación (CIDE), Valencia
Centro Mediterráneo de Investigaciones Marinas y Ambientales (CMIMA), Barcelona
Estación Biológica de Doñana (EBD), Sevilla
Estación Experimental de Zonas Áridas (EEZA), Almería
Instituto Andaluz de Ciencias de la Tierra (IACT), Granada

Instituto Botánico de Barcelona (IBB), Barcelona
Instituto de Acuicultura de Torre de la Sal (IATS), Castellón
Instituto de Biología Evolutiva (IBE), Barcelona
Instituto de Ciencias de la Tierra "Jaume Almera" (ICTJA), Barcelona
Instituto de Ciencias del Mar (ICM), Barcelona
Instituto de Ciencias Marinas de Andalucía (ICMAN), Cádiz
Instituto de Geología Económica (IGE), Madrid
Instituto de Investigación en Recursos Cinegéticos (IREC), sede Albacete
Instituto de Investigación en Recursos Cinegéticos (IREC), sede Ciudad Real
Instituto de Investigaciones Marinas (IIM), Vigo
Instituto de Recursos Naturales (IRN), Madrid
Instituto Mediterráneo de Estudios Avanzados (IMEDEA), Mallorca
Instituto Pirenaico de Ecología (IPE-H), Huesca
Instituto Pirenaico de Ecología (IPE-Z), Zaragoza
Museo Nacional de Ciencias Naturales (MNCN), Madrid
Real Jardín Botánico (RJB), Madrid
Unidad de Tecnología Marina (UTM), Barcelona

Área 4. Ciencias Agrarias

Centro de Ciencias Medioambientales (CCMA), Madrid
Centro de Edafología y Biología Aplicada del Segura (CEBAS),Murcia
Estación Agrícola Experimental (EAE), León
Estación Experimental Aula Dei (EEAD), Zaragoza
Estación Experimental del Zaidín (EEZ), Granada
Estación Experimental La Mayora (EELM), Málaga
Instituto de Agricultura Sostenible (IAS), Córdoba
Instituto de Agrobiotecnología (IdAB), Navarra
Instituto de Ciencias Agrarias (ICA), Madrid
Instituto de Ganadería de Montaña (IGM), León
Instituto de Investigaciones Agrobiológicas de Galicia (IIAG), Santiago de Compostela
Instituto de Recursos Naturales y Agrobiología (IRNAS), Sevilla
Instituto de Recursos Naturales y Agrobiología (IRNASA), Salamanca
Misión Biológica de Galicia (MBG), Pontevedra

Área 5. Ciencia y Tecnologías Físicas

Centro de Astrobiología (CAB), Madrid
Centro Nacional de Aceleradores (CNA), Sevilla
Centro Técnico de Informática (CTI), Madrid
Instituto de Acústica (IA), Madrid
Instituto de Astrofísica de Andalucía (IAA), Granada
Instituto de Astronomía y Geodesia (IAG), Madrid
Instituto de Automática Industrial (IAI), Madrid
Instituto de Ciencias del Espacio (ICE), Barcelona
Instituto de Ciencias Matemáticas (ICMAT), Madrid
Instituto de Estructura de la Materia (IEM), Madrid
Instituto de Física Aplicada (IFA) , Madrid
Instituto de Física Corpuscular (IFIC), Valencia
Instituto de Física de Cantabria (IFCA), Santander
Instituto de Física Fundamental (IFF), Madrid
Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), Palma de Mallorca
Instituto de Física Teórica (IFT), Madrid
Instituto de Investigación en Inteligencia Artificial (IIIA), Barcelona
Instituto de Matemáticas y Física Fundamental (IMAFF), Madrid
Instituto de Microelectrónica (IMM-CNM), Madrid
Instituto de Microelectrónica de Barcelona (IMB-CNM)
Instituto de Microelectrónica de Sevilla (IMSE-CNM)
Instituto de Óptica "Daza Valdés" (IO), Madrid
Instituto de Robótica e Informática Industrial (IRI), Barcelona
Laboratorio de Física de Sistemas Pequeños y Nanotecnología (FSP), Madrid

Laboratorio de Investigación en Tecnologías de la Combustión (LITEC), Zaragoza
Observatorio de Física Cósmica del Ebro (OE), Tarragona
Área 6. Ciencia y Tecnología de Materiales
Centro de Física de Materiales (CFM), San Sebastián
Centro de Investigación de Nanomateriales y Nanotecnología (CINN), Oviedo
Centro de Investigación en Nanociencia y Nanotecnología (CIN2), Barcelona
Centro Nacional de Investigaciones Metalúrgicas (CENIM), Madrid
Instituto de Cerámica y Vidrio (ICV), Madrid
Instituto de Ciencia de Materiales de Aragón (ICMA), Zaragoza
Instituto de Ciencia de Materiales de Barcelona (ICMAB)
Instituto de Ciencia de Materiales de Madrid (ICMM)
Instituto de Ciencia de Materiales de Sevilla (ICMS)
Instituto de Ciencia y Tecnología de Polímeros (ICTP), Madrid
Instituto de Ciencias de la Construcción "Eduardo Torroja" (IETcc), Madrid
Área 7. Ciencia y Tecnología de Alimentos
Instituto de Agroquímica y Tecnología de los Alimentos (IATA), Valencia
Instituto de la Grasa (IG), Sevilla
Instituto de Ciencia y Tecnología de Alimentos y Nutrición (ICTAN), Madrid
Instituto de Fermentaciones Industriales (IFI), Madrid
Instituto de las Ciencias de la Vid y el Vino (ICVV), Logroño
Instituto de Nutrición y Bromatología (INB), Madrid
Instituto de Productos Lácteos de Asturias (IPLA), Villaviciosa
Instituto del Frío (IF), Madrid
Área 8. Ciencia y Tecnologías Químicas
Centro de Investigación y Desarrollo (CID), Barcelona
Centro de investigaciones Científicas Isla de La Cartuja (CARTUJA), Sevilla
Centro de Química Orgánica "Manuel Lora Tamayo" (CENQUIOR), Madrid
Instituto de Carboquímica (ICB), Zaragoza
Instituto de Catálisis y Petroleoquímica (ICP), Madrid
Instituto de Diagnóstico Ambiental y Estudios del Agua (IDAEA), Barcelona
Instituto de Investigaciones Químicas (IIQ), Sevilla
Instituto de Investigaciones Químicas y Ambientales (IIQABD), Barcelona
Instituto de Productos Naturales y Agrobiología (IPNA), Santa Cruz de Tenerife
Instituto de Química Avanzada de Cataluña (IQAC) Barcelona
Instituto de Química Física "Rocasolano" (IQFR), Madrid
Instituto de Química Médica (IQM), Madrid
Instituto de Química Orgánica General (IQOG), Madrid
Instituto de Tecnología Química (ITQ), Valencia
Instituto Nacional del Carbón (INCAR), Oviedo