



VNiVERSiDAD
D SALAMANCA

CAMPUS OF INTERNATIONAL EXCELLENCE

Departamento de Estadística
Máster en Análisis Avanzado de Datos Multivariantes
Trabajo Fin de Máster

Revisión del Modelo Unfolding, Un modelo geométrico para datos de elección y preferencia.

Autora: VANESSA M. VERA LÓPEZ
Tutora: ANA MARÍA MARTÍN CASADO

2013

UNIVERSIDAD DE SALAMANCA
Departamento de Estadística
Máster en Análisis Avanzado de Datos Multivariantes
Trabajo Fin de Máster

Revisión del Modelo Unfolding, Un modelo geométrico para datos de elección y preferencia.

Autora: VANESSA M. VERA LÓPEZ
Tutora: ANA MARÍA MARTÍN CASADO

2013



Dpto. de Estadística
Universidad de Salamanca

DRA. ANA MARÍA MARTÍN CASADO

Profesora Titular del Departamento de Estadística de la Universidad de Salamanca

CERTIFICA que **D.^a Vanessa María Vera López** ha realizado en la Universidad de Salamanca, bajo su supervisión y dirección, el trabajo que para optar al título de Máster en Análisis Avanzado de Datos Multivariantes, presenta con el título ***Revisión del Modelo Unfolding, un modelo geométrico para datos de elección y preferencia***, autorizando expresamente su lectura y defensa.

Y para que conste, firma el presente certificado en Salamanca a 26 de agosto de 2013.

Ana M^a Martín Casado

Aquel que duda y no investiga, se torna no sólo infeliz, sino también injusto.

Blaise Pascal (1623-1662) Matemático, físico, filósofo y escritor francés.

Pelearé hasta el último segundo y mi epitafio será: No estoy de acuerdo. Joaquín Sabina Poeta y Cantautor.

Dedicatoria

A mis padres, Justo y Mercedes, por ofrecerme todas las oportunidades que no tuvisteis, quererme, educarme, creer en mí y en mi afición desde niña por las Matemáticas que se han convertido en mi modo de vida. Por apoyarme en los malos momentos que también los hubo, siempre supisteis que este ratón de biblioteca llegaría allá dónde se propusiera con tesón, esfuerzo y trabajo. Gracias a vosotros he podido llegar hasta aquí.

A Alfredo, por quererme tanto y tan bien, por creer en mí y en mi capacidad...por sufrir junto a mí en este trabajo...gran parte de esto te lo debo a tí.

A Orión, nuestra pequeña estrellita de Febrero.

A mi familia.

A la memoria de mi padre, se fue antes de verme licenciada, no me imagino lo orgulloso que estaría de saber que he llegado hasta aquí. Tío Joaquín, Abuela Trini y Abuela Benita, siempre estáis en mi pensamiento.

Agradecimientos

A Alfredo, Elina, tía Puri por la ayuda.

A mi madre por soportarme siempre, en fin por TODO, Abraham, tía Mari, Ramiro y Guille.

A todos mis Amigos, por aceptarme así como soy...

A Pablo Rebollo por confiar en mí y darme la oportunidad de trabajar en BAP. También a todos los que confiaron en mí en algún momento, de todos mis trabajos he aprendido algo siempre.

A todo el Departamento de Estadística de la Universidad de Salamanca, por ese enorme esfuerzo en sacar el Máster adelante. En especial quería agradecer a Puri (Dra. Galindo) por su enorme empatía y apoyo recibido durante todo el curso y también por sus riñas siempre justificadas y a la Dra. Ana M^a Martín por su ayuda, consejos y conocimientos en la realización de este trabajo.

Índice

ÍNDICE	3
ÍNDICE DE TABLAS.....	4
ÍNDICE DE FIGURAS	4
RESUMEN.....	5
INTRODUCCIÓN.....	6
OBJETIVOS	10
MATERIAL Y MÉTODOS	11
ESCALAMIENTO MULTIDIMENSIONAL.....	11
<i>Modelos de MDS.....</i>	<i>13</i>
<i>Errores, Funciones de Pérdida y Stress.....</i>	<i>15</i>
UNFOLDING: MODELO PARA DATOS DE ELECCIÓN O PREFERENCIA.	18
<i>Una aproximación inicial: Modelo de Punto-Ideal e Isopreferencias.....</i>	<i>19</i>
<i>Soluciones Degeneradas o Triviales.</i>	<i>20</i>
<i>Configuración de Puntos en el Espacio Común (Regiones isótonas e indeterminaciones).....</i>	<i>26</i>
<i>Otras consideraciones en Unfolding Multidimensional</i>	<i>27</i>
<i>Soluciones Triviales, ¿cómo resolverlas?: Tres enfoques distintos.....</i>	<i>28</i>
Ajuste o adaptación de los datos	28
Ajuste de la transformación (<i>Adjusting the Transformation</i>).....	31
Ajustes en la Función de Pérdida (<i>Adjustments to the Loss Function</i>)	32
<i>Modelos de Unfolding.....</i>	<i>36</i>
<i>Revisión del Software en Unfolding</i>	<i>41</i>
MINISSA.....	41
KYST.....	42
ALSCAL.....	42
NEWMDSX.....	43
SMACOF-III	44
GENFOLD-2.....	44
NEWFOLD.....	45
PREFSCAL.....	45
<i>Un paso más: Unfolding de Tres Vías (Three-Way Unfolding).....</i>	<i>45</i>
Modelo Ponderado de Tres Vías (<i>The Weighted Least Squares Algorithm for Three-Way Unfolding</i>)	48
Algoritmo de Mínimos Cuadrados Ponderados para Unfolding de Tres Vías.....	50
<i>Desarrollos Recientes: Unfolding de Tres Vías y Análisis Cluster</i>	<i>55</i>
Modelo Cluster en Unfolding de Tres Vías	57
Algoritmo de estimación	58
RESULTADOS.....	65
<i>Aplicaciones del Unfolding de Tres Vías.....</i>	<i>65</i>
Aplicación del modelo ponderado de tres vías de DeSarbo y Carroll (1985)	66
Aplicación del Modelo de Tres Vías y Cluster de DeSarbo et al. (2009)	71
<i>Implementación del Algoritmo Weighted Least Squares Algorithm for Three-Way Unfolding de 1985 en Código MATLAB.....</i>	<i>76</i>
CONCLUSIONES.....	89
BIBLIOGRAFÍA.....	91

Índice de Tablas

TABLA 1 .FACTORES INDEPENDIENTES Y NIVELES DEFINIDOS PARA EL ANÁLISIS MONTE CARLO EN UNFOLDING DE TRES VÍAS (FUENTE DESARBO Y CARROLL, 1985).....	54
TABLA 2 .FACTORES INDEPENDIENTES Y NIVELES DEFINIDOS PARA EL ANÁLISIS MONTE CARLO EN UNFOLDING DE 3 VÍAS Y CLUSTER (FUENTE DESARBO ET AL., 2009).....	62
TABLA 3. PRODUCTOS EN EJEMPLO DE UNFOLDING DE TRES VÍAS Y CLUSTER (FUENTE DESARBO ET AL., 2009)	72
TABLA 4. ATRIBUTOS EN EJEMPLO DE UNFOLDING DE TRES VÍAS Y CLUSTER (FUENTE DESARBO ET AL., 2009)	72
TABLA 5. TABLA DE RESULTADOS DE SIMULACIONES ALGORITMO 3 VÍAS DESARBO (1985).....	78

Índice de Figuras

FIG. 1 EJEMPLO DE CURVA DE RESPUESTA UNIMODAL.....	7
FIG. 2. EJEMPLO DE DIAGRAMA DE SHEPARD.....	17
FIG. 3. MATRIZ DE PREFERENCIAS	19
FIG. 4. UN EJEMPLO DE SOLUCIÓN UNFOLDING (GREEN Y RAO, 1972).....	20
FIG. 5. UN EJEMPLO DE SOLUCIÓN TRIVIAL EQUIDISTANTE	22
FIG. 6. UN EJEMPLO DE SOLUCIÓN DEGENERADA PARA UNFOLDING ORDINAL INCONDICIONAL Y DIAGRAMA DE SHEPARD ASOCIADO CON EL STRESS-1.	23
FIG. 7. UN EJEMPLO DE SOLUCIÓN DEGENERADA PARA UNFOLDING ORDINAL INCONDICIONAL Y DIAGRAMA DE SHEPARD ASOCIADO CON EL STRESS-2.	24
FIG. 8. DIAGRAMA DE SHEPARD PARA UNFOLDING INCONDICIONAL DE INTERVALO CON STRESS-1.	25
FIG. 9. SOLUCIÓN UNFOLDING INCONDICIONAL DE INTERVALO Y DIAGRAMA DE SHEPARD CON STRESS-2.	26
FIG. 10. EJEMPLO DE REGIONES ISÓTONAS CON 4 ESTÍMULOS EN UN ESPACIO DE 2 DIMENSIONES (Cox, 2000)	27
FIG. 11. REPRESENTACIÓN GRÁFICA DE LA FUNCIÓN DE PESOS w_{ij} (BORG Y GROENEN, 2005).....	33
FIG. 12. EJEMPLO DE SOLUCIÓN UNFOLDING LINEAL MEDIANTE PREFSCAL Y DIAGRAMA DE SHEPARD	36
FIG. 13. EJEMPLO DE SOLUCIÓN UNFOLDING CONDICIONAL MEDIANTE PREFSCAL Y DIAGRAMA DE SHEPARD	36
FIG. 14. EJEMPLO DE MODELO VECTORIAL: L1 Y L2 SON DOS INDIVIDUOS (BORG Y GROENEN, 2005) .	39
FIG. 15. REPRESENTACIÓN DEL ESPACIO CONJUNTO EJEMPLO 1 (DESARBO Y CARROLL, 1985)	67
FIG. 16. PESOS OBTENIDOS EJEMPLO1 (DESARBO Y CARROLL, 1985).....	68
FIG. 17. ESPACIO CONJUNTO EN 1 DIMENSIÓN EJEMPLO 2 (DESARBO ET AL., 1985)	71
FIG. 18. ESPACIO CONJUNTO SIN CLUSTERS (PREFERENCIA GENERAL) (DESARBO ET AL., 2009).....	73
FIG. 19. ESPACIO CONJUNTO CLUSTER 1 (DESARBO ET AL., 2009)	74
FIG. 20. ESPACIO CONJUNTO CLUSTER 2 (DESARBO ET AL., 2009)	75
FIG. 21. ESPACIO CONJUNTO CLUSTER 3 (DESARBO ET AL., 2009)	75
FIG. 22. EJEMPLO DE REPRESENTACIÓN CONJUNTA CON PUNTOS FIJOS EN LA SIMULACIÓN MATLAB ALGORITMO 3 VÍAS DESARBO (1985)	78
FIG. 23. EJEMPLO DE REPRESENTACIÓN CONJUNTA CON PUNTOS FLOTANTES EN LA SIMULACIÓN MATLAB ALGORITMO 3 VÍAS DESARBO (1985).....	79

Resumen

Este trabajo toma como punto de partida la revisión del modelo Unfolding realizada por Ingwer Borg y Patrick Groenen en su libro '*Modern Multidimensional Scaling*' publicado en 2005, en el cual se realiza un análisis profundo del Escalamiento Multidimensional (en adelante MDS, por sus siglas en inglés, *multidimensional scaling*), sus aplicaciones a distintos campos de la ciencia aparte de la Psicología y sus diferentes modelos. El análisis incluye aspectos técnicos y matemáticos como son el Álgebra Lineal, la Geometría, las funciones empleadas para la transformación de los datos originales, los algoritmos utilizados para encontrar la representación más fiel de estos datos y el tratamiento de los errores estadísticos en cada uno de los modelos.

Borg y Groenen dedican tres capítulos a revisar y exponer los aspectos más relevantes del Unfolding como un caso especial del MDS enfocado a datos de preferencia: las dificultades que cualquier investigador podría encontrar en el tratamiento de los datos de entrada, la posible aparición de soluciones triviales y degeneradas, el tratamiento de datos ausentes en la matriz de preferencias, la búsqueda de funciones que transformen de forma rigurosa los datos originales en distancias y un correcto ajuste en la función de pérdida.

Se realiza además una revisión cronológica del software disponible para el cálculo de soluciones y se completa con los últimos avances de la técnica dirigidos hacia el Unfolding de Tres Vías y su fusión con el Análisis Cluster.

Para finalizar se exponen distintas aplicaciones prácticas de esta técnica realizadas en diversos campos.

Palabras Clave: MDS, Unfolding, Modelos de Preferencia, Soluciones Triviales y Degeneradas, Unfolding de Tres Vías.

Introducción

Según Cuadras (2012), el análisis multivariante (AM) es la parte de la estadística y del análisis de datos que estudia, analiza, representa e interpreta los datos que resultan de observar un número $p > 1$ de variables estadísticas sobre una muestra de n individuos. Las variables observables son homogéneas y correlacionadas, sin que alguna predomine sobre las demás. La información estadística en AM es de carácter multidimensional, por lo tanto la geometría, el cálculo matricial y las distribuciones multivariantes juegan un papel fundamental pues la estadística tiene como soporte teórico diversas ramas de la matemática como son el análisis, la probabilidad, el álgebra, la geometría, etc. (Cuadras,1985).

La información multivariante normalmente, se puede almacenar en una matriz de datos pero, a menudo, en AM la información de entrada consiste en matrices de distancias o similitudes, que miden el grado de discrepancia entre los individuos; en concreto, este tipo de análisis o técnica es el conocido como escalamiento multidimensional (MDS), representación multidimensional mediante escalas o análisis de proximidades (Martín-Casado y Galindo, 1994). Su objetivo es encontrar el número de dimensiones adecuado a los datos y localizar los estímulos (u objetos percibidos) sobre cada dimensión, calculando sus coordenadas, de modo que las distancias entre ellos se adapten tan estrechamente como sea posible a las proximidades entre los objetos o a transformaciones de ellas.

El MDS representa los objetos que se juzgan similares por puntos cercanos en la representación espacial resultante, mientras que los objetos juzgados disimilares se representan por puntos alejados. Además, proporciona un espacio que descubre relaciones relevantes a los observadores y/o la estructura subyacente en los datos.

Como se ha comentado antes, además del Álgebra, dentro del MDS es fundamental el conjunto de fundamentos y propiedades de la Geometría como son los espacios métricos (espacios topológicos cuya topología viene inducida a través de una distancia).

Otro de los hitos relevantes en el desarrollo del AM y del MDS ha sido la evolución de la Informática, que ha hecho posible desarrollar e implementar programas estadísticos necesarios para la manipulación algebraica de las matrices de datos.

El Modelo Unfolding surge como una técnica derivada del MDS, con la que se pretende situar en un mismo espacio perceptual los conjunto de individuos y de estímulos en base a ciertos juicios de preferencia.

De acuerdo a lo sugerido por Coombs y Avrunin (1977, 1988), la Teoría de Conflictos, Intereses o Preferencias surge como solución a determinados conflictos incompatibles. Bajo esas circunstancias, la preferencia de un individuo puede describirse mediante una función o curva de respuesta unimodal (SPF o *Single Peaked Function*).

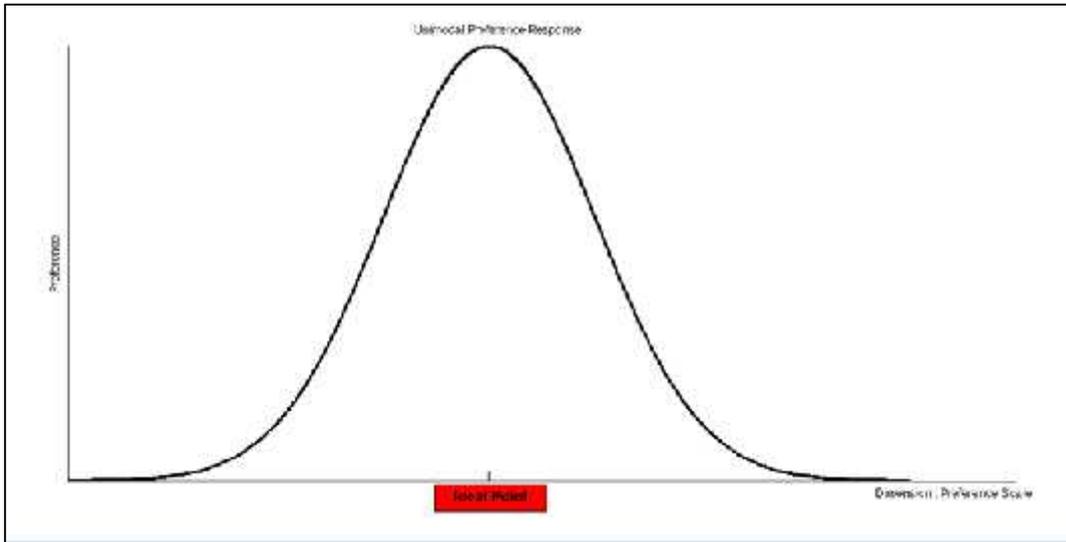


Fig. 1 Ejemplo de Curva de Respuesta Unimodal

La curva de respuesta unimodal permite una búsqueda o proceso que converja a la elección óptima, por ejemplo cuando uno está en la ducha es relativamente fácil graduar la temperatura del agua para conseguir la elección personal óptima en forma de satisfacción personal.

Además de los conflictos intrapersonales (los que surgen en la vida cotidiana a un mismo individuo), hay que tener en cuenta los conflictos interpersonales, en los que un grupo de individuos debe de elegir entre varias alternativas conforme a sus propios intereses, que se conoce como PRT o *Person Response Theory*. Bajo este supuesto:

- Las curvas de respuesta unimodal para cada individuo se toman como parámetros de posición o puntos ideales.
- Los estímulos son valorados en una escala.
- Los datos están condicionados por cada individuo (los individuos otorgan puntuaciones a una serie de estímulos en base a la preferencia que tengan hacia ellos).
- Dada la naturaleza del problema, el análisis debe ser multidimensional.

La medida y el conocimiento de las diferencias entre individuos necesita de una explicación y además es necesario el poder anticiparse a ellos a través de métodos estadísticos adecuados.

Según Busing (2010), el Unfolding nace en el ámbito de la Psicología y las Ciencias Sociales. Coombs (discípulo de Thurstone) introduce el concepto cuando publica un artículo en la revista *Psychological Review* en el cual demuestra que un simple ranking de preferencias contiene información métrica (Coombs, 1950). Tras este artículo, el concepto de Unfolding continúa desarrollándose en base a las ideas de Thurstone (método de las comparaciones pareadas) y de Guttman (1944, 1946), quien demostró mediante la escala Guttman que los ítems que miden cierta actitud poseen intensidad y pueden ser escalables en un sentido de jerarquía y orden.

Así, Coombs desarrolla un nuevo tipo de escala y define la escala conjunta J, en la cual tanto individuos como estímulos ocupan posiciones fijas en un espacio común, la posición de un individuo representa el *ideal* en el sentido de que el individuo estará más cercano al estímulo hacia el cual ha mostrado mayor preferencia.

El término Unfolding (“desdoblado”, en español) fue elegido por Coombs por esta sencilla razón: si se imagina la representación conjunta en la escala J de Coombs dibujada en un pañuelo, se sujeta el punto que representa un individuo concreto con dos dedos, este punto queda en la parte superior y los estímulos quedan representados sobre una “curva” de manera que los estímulos menos preferidos estarán más lejos del individuo, es decir, si se observa la configuración del pañuelo doblado, se puede adivinar el orden de preferencia que ha otorgado el individuo a los estímulos. Además, el orden de los estímulos en la escala J de Coombs es la escala I del individuo (puntuaciones otorgadas a los estímulos) y el Unfolding o desdoblamiento es la opción inversa: los órdenes de la preferencia de cada individuo (escala I conocida) conforman los datos a representar en la escala conjunta J desconocida.

Esta idea pronto fue extendida por Bennett y Hays (1960,1961) al caso multidimensional y en ambos artículos se dio a conocer el Unfolding Multidimensional, centrándose en el problema de determinar la mínima dimensión necesaria para la representación de los datos y la obtención dicha configuración. Pero pronto surgieron los primeros inconvenientes: los métodos que hasta ese momento se venían empleado para obtener las soluciones no eran manejables para datos ordinales, pues producían soluciones no métricas (los individuos no se representaban mediante puntos fijos sino

que podían yacer en regiones isótonas¹) y además se añadía la dificultad de la no existencia de programas informáticos adecuados.

Desde entonces, se establece una búsqueda incansable para solucionar las dificultades que surgen motivadas básicamente por el tratamiento y la medida de los datos, la necesidad de encontrar soluciones métricas para datos no métricos, la búsqueda de una medida de ajuste para la configuración conjunta que minimice la diferencia entre disparidades y las distancias, el desarrollo de algoritmos que permitan obtener soluciones válidas y ajustadas a los datos iniciales, además de la necesidad de la informática para facilitar el cálculo de dichos algoritmos.

Es por este motivo que el Unfolding como técnica que deriva del MDS y como parte del AM se revela como un método completo desde el punto de vista de la estadística, porque combina aspectos matemáticos como la manipulación algebraica de las matrices de datos, la geometría embebida dentro de la solución y la necesidad del desarrollo de programas informáticos para la obtención de soluciones a través de los algoritmos que, en gran medida, están basados en Métodos Numéricos que provienen del Análisis Matemático.

Además, con el desarrollo en el análisis de matrices de tres entradas y su aplicación en el MDS y como respuesta a la necesidad de entender las diferencias entre individuos no solo en una determinada situación sino en determinados contextos diferentes, surge el Unfolding de tres vías, ya que si en una determinada situación un individuo puede ordenar una colección de estímulos según un rango de preferencias y el Unfolding proporciona una solución óptima al problema, desde luego cabe la posibilidad de poder generar soluciones en dos o más situaciones.

Para finalizar se puede resaltar el hecho de la poca repercusión de esta técnica en nuestro país a pesar de las interesantes aplicaciones que tiene, fundamentalmente en Psicología, Ciencias Sociales (Sociología, Política), Marketing (Investigación de Mercados) y pequeñas aportaciones en otras disciplinas como se verá más adelante.

Este trabajo por tanto constituye una revisión bibliográfica del desarrollo y avances del Unfolding, abarcando referencias desde la primera contribución teórica hasta los desarrollos más recientes que llevan al modelo de tres entradas y a la combinación con otras técnicas de AM.

¹La definición de región isótónica se muestra en el capítulo siguiente.

Objetivos

El objetivo principal de este trabajo es revisar la técnica de Unfolding Multidimensional (MDU, por sus siglas en inglés), desde la introducción de la idea, sugerida por Coombs a principio de los años 50, hasta el momento actual.

Además, mediante la lectura y análisis de la literatura publicada referente a este tema se pretende abordar los siguientes objetivos:

- Revisar el Álgebra y la Geometría subyacente.
- Realizar una aproximación al problema de las soluciones triviales o degeneradas y las técnicas propuestas por diversos autores para solventarlas.
- Revisar los distintos modelos de Unfolding propuestos (Condicionados, Internos/Externos, etc.)
- Profundizar en el Unfolding de tres vías, estudiando y analizando la bibliografía más relevante del tema.
- Revisar los algoritmos empleados en Unfolding así como los diversos programas informáticos que permiten su implementación.
- Buscar y exponer distintas aplicaciones prácticas de esta técnica realizadas en diversos campos.
- Desarrollar e implementar uno de los algoritmos revisados en este trabajo en Matlab.

Material y Métodos

Escalamiento Multidimensional

Uno de los problemas más interesantes en muchas disciplinas se plantea cuando se necesita entender las relaciones entre objetos y se desconocen las dimensiones subyacentes en la comparación de los mismos, especialmente en aquellas situaciones en las que la información disponible se refiere exclusivamente a la similaridad o disimilaridad entre los objetos que son motivo de estudio. El Escalamiento Multidimensional, en adelante (MDS) fue originariamente concebido como un procedimiento que permite la construcción de una configuración de puntos en un espacio de dimensión reducida, a partir de la información proporcionada por las medidas de proximidad (similaridades o disimilaridades) entre cada par de objetos a representar.

El MDS es, por tanto, una técnica estadística clasificada dentro del Análisis Multivariante cuyos orígenes se remontan a finales del siglo XIX, que permite la representación de una colección de objetos partiendo de ciertas medidas de similaridad (o disimilaridad) entre ellos. Borg y Groenen (2005) señalan cuatro propósitos básicos de esta técnica:

- i. Reproducir medidas de similaridad o disimilaridad mediante distancias en un espacio de baja dimensión para facilitar su interpretación.
- ii. Probar si ciertas características que se pueden usar para distinguir objetos son las responsables de las diferencias percibidas entre ellos.
- iii. Identificar dimensiones en base a juicios de similaridad.
- iv. Explicar juicios de disimilaridad creando una regla que refleje el tipo de distancia que permite diferenciar los objetos.

Este método comprende un conjunto de procedimientos y técnicas que tratan de representar y explicar la configuración obtenida en un entorno multidimensional. Así por ejemplo, dada una matriz de correlaciones entre diversas variables, el MDS permite representar las variables como puntos, de forma que dos puntos cualesquiera se encontrarán tanto más próximos entre sí cuanto más correlacionadas estén las variables a las que representan. Si las distancias interpuntuales se corresponden suficientemente bien con las correlaciones, es posible conseguir una representación que ponga de manifiesto la estructura intrínseca existente que, de otro modo, podría

permanecer oculta al investigador puesto que en general resulta mucho más difícil observar una tabla de coeficientes de correlación que una gráfica en un plano.

El siguiente ejemplo permite entender de forma rápida la técnica de MDS:

Se supone que se está interesado en el estudio de ciudades respecto a su posición geográfica. Si se dispone de un mapa donde se encuentran representadas las ciudades la construcción de la correspondiente matriz de distancias entre las ciudades a partir de dicho mapa se reduce a una cuestión gráfica elemental. Supóngase, por el contrario, que se dispone de una matriz cuadrada formada por las distancias lineales entre cada pareja de ciudades y se pretende reconstruir el mapa partiendo de dicha matriz. Cuando la matriz de distancias es euclídea, este problema **tiene solución exacta** en una determinada dimensión, como se recoge en el siguiente resultado (teorema) (Young y Householder, 1938):

Sea $D_{n \times n} = (\delta_{ij})$ una matriz de distancias entre n puntos y sea $B_{n \times n}$ la matriz dada por $B = H\Delta H$ con $H_{n \times n}$, cuya expresión es $H = I - n^{-1}1_n 1_n^t$ con $\Delta_{n \times n}$ la matriz cuyos datos vienen dados por $\Delta_{ij} = -\frac{1}{2}\delta_{ij}^2$. Entonces D es una matriz de distancias euclídeas si y sólo si B es semidefinida positiva². Además se tiene que:

- i. Si $D_{n \times n}$ es la matriz de distancias euclídeas para una configuración de puntos en un espacio de dimensión K dada por $Z_{(n \times K)} = (z_1, \dots, z_n)^t$, entonces $B = (HZ)(HZ)^t$, es decir, los elementos de B se expresan como $b_{rs} = (z_r - \bar{z})^t (z_s - \bar{z}) \forall r, s: 1 \dots n$, de donde se desprende que $B \geq 0$ y B es la matriz centrada de productos escalares.
- ii. Recíprocamente, si B es semidefinida positiva de rango K , entonces puede construirse una configuración asociada a B de la siguiente forma: Sean $\lambda_1 > \lambda_2 > \dots > \lambda_k$ los k valores propios positivos y $x_{(1)}, \dots, x_{(k)}$ sus k vectores propios normalizados según la condición $x_{(i)}^t x_{(i)} = \lambda_i \forall i: 1 \dots k$. Entonces las distancias euclídeas entre los puntos de \mathbb{R}^k de coordenadas $x_r = (x_{r1}, \dots, x_{rk})^t$ donde x_r es la r -ésima fila de la matriz $X = (x_{(1)}, \dots, x_{(r)})$, están dadas por la matriz de distancias D . Además esta configuración está centrada ($\bar{x}=0$) y B es la matriz de productos escalares de esta configuración.

² Una matriz simétrica A es **semidefinida positiva** si $x^t Ax = 0$ para todos los posibles vectores x (excepto $x=0$). Los valores propios de una matriz semidefinida positiva son todos positivos o nulos.

Este teorema, cuya demostración puede consultarse en Mardia (1980), indica que existe solución única para distancias euclídeas en un espacio de dimensión $K = \text{rg}(B)$ salvo isometrías, donde K alcanza como máximo el valor $(n-1)$.

Modelos de MDS

Son muchos los algoritmos que se han desarrollado para obtener la configuración que mejor se ajuste a las disimilaridades de partida. Estos algoritmos están fundamentados en modelos espaciales de naturaleza geométrica en los que las disimilaridades se suponen relacionadas mediante una cierta función con las distancias interpuntuales de la configuración resultante. Cabe distinguir entre los modelos que utilizan disimilaridades medidas en escala de intervalo o de razón, que son los modelos métricos, y los que utilizan disimilaridades medidas en escala ordinal, que son los modelos no métricos.

En ese sentido, es necesario conocer y distinguir el tipo de medida o escala de los datos. Debe recordarse que:

- La escala nominal sólo permite establecer relaciones de igualdad y desigualdad entre los datos.
- La escala ordinal sólo permite establecer una jerarquía entre los datos en un sentido de más a menos o viceversa.
- La escala de intervalo permite ordenar no sólo los datos sino también las diferencias entre ellos.
- La escala de razón permite establecer relaciones de proporcionalidad entre los datos.

Además, es adecuado conocer la condicionalidad de la medida, ya que si todos los sujetos utilizan del mismo modo la escala de medida y llevan a cabo sus estimaciones siguiendo los mismos criterios los datos son **incondicionales**. Sin embargo, si los datos no son comparables entre sí son datos **condicionales**, pudiéndose distinguir entre:

- Condicionalidad por matriz, cuando los datos son comparables dentro de una matriz pero no son comparables de una matriz a otra (análisis de tres vías).
- Condicionalidad por fila, cuando no se puede realizar la comparación de una fila a otra de la matriz. En este caso, se parte de una matriz rectangular de preferencias, donde cada fila representa un sujeto y se supone que los sujetos establecen sus preferencias en base a criterios diferentes.

Para describir los diferentes modelos de MDS se puede hacer referencia en principio y sin pérdida de generalidad, al tipo de datos más usual que consiste en una matriz cuadrada de disimilaridades δ_{ij} de dos vías y un modo. Se puede suponer, por tanto, un conjunto de n objetos y una matriz de disimilaridades entre los pares de objetos. El objetivo del MDS es encontrar una configuración de puntos en un espacio de dimensión K de forma que las distancias d_{ij} se adapten tanto como sea posible a las disimilaridades δ_{ij} , o a una transformación de ellas:

$$f(\delta_{ij}) \approx d_{ij}$$

La elección de la función o transformación f especificará el tipo de modelo MDS:

- **Modelos de MDS Métrico:** Son los modelos de MDS en los que el tipo de relación funcional que se establece entre las disimilaridades y las distancias tiene en cuenta, además del orden, el valor intrínseco de las disimilaridades. Estos modelos suelen estar asociados a datos medidos en escalas de intervalo y de razón y tratan de encontrar un conjunto de puntos en un espacio de forma que cada punto represente uno de los objetos y las distancias entre los puntos sean aproximadamente iguales cuantitativamente a las correspondientes disimilaridades transformadas mediante la relación $f(\delta_{ij}) \approx d_{ij}$ donde f es una función continua, paramétrica y monótona. Dentro de estos modelos pueden destacarse principalmente los siguientes:

- MDS clásico o de Torgerson : $f(\delta_{ij}) = \delta_{ij}$
- MDS para datos medidos en escala de razón : $f(\delta_{ij}) = b\delta_{ij}$
- MDS para datos medidos en escala de intervalo : $f(\delta_{ij}) = a + b\delta_{ij}$

La función f de los tres modelos anteriores es una función lineal, pero esto no tiene por qué ser así. Se puede elegir cualquier tipo de función. Otros modelos de MDS métrico no lineales son:

- MDS logarítmico : $f(\delta_{ij}) = b \cdot \log(\delta_{ij})$ ó $f(\delta_{ij}) = a + b \cdot \log(\delta_{ij})$
- MDS exponencial : $f(\delta_{ij}) = a + b e^{\delta_{ij}}$
- MDS cuadrático : $f(\delta_{ij}) = a + b\delta_{ij} + c\delta_{ij}^2$
- MDS inverso : $f(\delta_{ij}) = a + b/\delta_{ij}$
- MDS recíproco : $f(\delta_{ij}) = b(1 - \delta_{ij})$

- **Modelos de MDS No Métrico:** Si se abandona la naturaleza métrica de la transformación de las disimilaridades por una relación puramente ordinal, se obtienen modelos de MDS no métricos. La transformación f en este caso puede ser arbitraria y solo obedece a la restricción monótona³, $\delta_{ij} < \delta_{rs} \rightarrow f(\delta_{ij}) \leq f(\delta_{rs})$; además, siempre hace uso de datos medidos en escala ordinal o nominal. Por tanto, solo las ordenaciones de las disimilaridades deben preservarse por la transformación y de ahí el término no métrico.

Los modelos no métricos son más flexibles que los métricos, en la medida en que no suponen ningún tipo de relación específica entre la distancia calculada y la medida de proximidad. Sin embargo, los resultados de los procedimientos métricos y no métricos aplicados a los mismos datos son a menudo muy similares. La razón de esta aparente paradoja es que, a pesar de la pérdida de información que supone transformar datos métricos en datos no métricos, la representación espacial impone grandes restricciones a las soluciones posibles.

Errores, Funciones de Pérdida y Stress

Tanto en los modelos métricos como en los no métricos se requiere que las proximidades, o transformaciones de ellas, sean representadas exactamente por sus correspondientes distancias. No obstante, en la práctica, la presencia de “ruido” en las proximidades empíricas hace que la relación de “igualdad” se relaje y sustituya por la de “aproximadamente igual”, de modo que no resulta necesaria la representación exacta sino que basta una buena aproximación de la solución. Así, una medida de la bondad de ajuste debe considerar el error total cometido en la aproximación.

En la definición de tal medida se emplea el concepto de error (cuadrático) de representación:

$$e_{ij}^2 = [f(\delta_{ij}) - d_{ij}]^2$$

Sumando todos los e_{ij}^2 , se obtiene el Stress bruto, una medida de “maldad” de ajuste para la configuración X obtenida:

³Una función entre conjuntos ordenados se dice **monótona** (o **isótona**) si conserva el orden dado.

$$\sigma_r(\mathbf{X}) = \sqrt{\sum_{i,j} e_{ij}^2} = \sqrt{\sum_{i,j} [f(\delta_{ij}) - d_{ij}(\mathbf{X})]^2} \quad (1)$$

Dado que en la práctica se pueden encontrar valores ausentes o desconocidos para δ_{ij} que no imponen restricción alguna en las distancias en el conjunto, se introducen los pesos w_{ij} cuyo valor será 1 o mayor que 1 si δ_{ij} es conocido y 0 si δ_{ij} no se conoce. Con los pesos, la expresión del Stress bruto queda:

$$\sigma_r(\mathbf{X}) = \sqrt{\sum_{i,j} w_{ij} [f(\delta_{ij}) - d_{ij}(\mathbf{X})]^2} \quad (2)$$

La raíz cuadrada de ese valor, normalizado por la suma de las distancias al cuadrado (para evitar la dependencia de las escalas de medida), es lo que se denomina Stress-1 (Kruskal, 1964a):

$$\sigma_1(\mathbf{X}) = \sqrt{\frac{\sum_{i,j} w_{ij} [f(\delta_{ij}) - d_{ij}(\mathbf{X})]^2}{\sum_{i,j} d_{ij}^2(\mathbf{X})}} \quad (3)$$

Minimizar el Stress-1 requiere encontrar una configuración óptima \mathbf{X} en una dimensión dada. Además, si f es una función paramétrica⁴, los valores de los parámetros también deberán ser estimados. Habitualmente la manera de hacerlo es mediante la regresión de las distancias interpuntuales para \mathbf{X} sobre las proximidades: una regresión lineal en MDS con datos medidos en escala de intervalo, una regresión monótona en MDS con datos medidos en escala ordinal. Los valores de las distancias estimados mediante la regresión son las disparidades o “distancias aproximadas”, denominadas también por algunos autores (Kruskal, 1977; Heiser, 1990) como pseudo-distancias y denotadas como \hat{d}_{ij} .

Para valorar el grado de ajuste de la configuración obtenida se puede usar un gráfico llamado diagrama de Shepard (Fig. 2), que no es más que un diagrama de dispersión que representa a las proximidades frente a las disparidades o distancias aproximadas.

⁴En matemáticas, una **ecuación paramétrica** permite representar una o varias curvas o superficies en el plano o en el espacio, mediante valores arbitrarios o mediante una constante, llamada **parámetro**.

Permite detectar *outliers* y, además, es particularmente útil en MDS ordinal para identificar la forma aproximada de la función monótona desconocida.

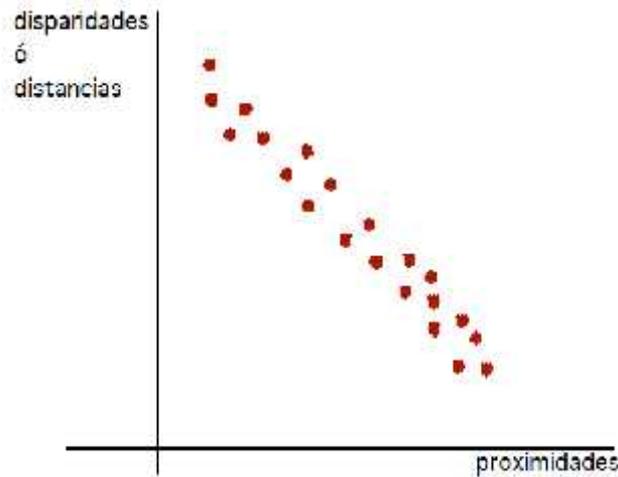


Fig. 2. Ejemplo de Diagrama de Shepard

Los programas informáticos que realizan escalamiento multidimensional efectúan el ajuste tanto de los modelos métricos como de los no métricos en forma iterativa mediante los siguientes pasos:

- 1º. Seleccionan una configuración inicial de puntos en una determinada dimensión.
- 2º. Calculan las distancias interpuntuales de la configuración de puntos del paso anterior y realizan una regresión de estas distancias sobre las proximidades, obteniendo las disparidades.
- 3º. Determinan una medida de “maldad” de ajuste. Las que se suelen usar son el Stress de Kruskal y el S-Stress de Takane, Young y DeLeeuw, definidos de la siguiente manera:

$$\text{Stress} = \sqrt{\frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}} \quad (4)$$

$$\text{S - Stress} = \sqrt{\frac{\sum_{i < j} (d_{ij}^2 - \hat{d}_{ij}^2)^2}{\sum_{i < j} d_{ij}^4}} \quad (5)$$

donde d_{ij} son las distancias interpuntuales de la configuración de puntos y \hat{d}_{ij} son las disparidades.

4º. Si la medida de ajuste no llega a un valor límite adecuado, buscan una nueva configuración de puntos para la que la medida de ajuste sea menor. Para ello

determinan las direcciones en las cuales puede obtenerse una mejora en el ajuste y a continuación mueven los puntos de la configuración en esas direcciones mediante pequeños incrementos. Por ejemplo, el algoritmo ALSCAL, implementado en el programa SPSS, actualiza las coordenadas de la configuración de puntos mediante el método del gradiente o método del mayor descenso.

5°. Una vez alcanzada la medida de ajuste satisfactoria, se reduce en uno el número de dimensiones de la solución y se repite el proceso hasta obtener una solución en el menor número posible de dimensiones con una medida aceptable de ajuste.

Unfolding: Modelo para datos de Elección o Preferencia.

El modelo de escalamiento “desdoblado” o Unfolding es un caso particular del MDS expuesto anteriormente, que utiliza como datos de entrada puntuaciones de preferencia dadas por un grupo de sujetos a un conjunto de estímulos y representa conjuntamente ambas entidades (sujetos y estímulos).

Originalmente fue desarrollado por Coombs (1964) para analizar datos de elección o preferencia y en este modelo se puede asumir que existe una relación de proximidad entre los elementos de dos conjuntos distintos de modo que cada elemento i (con $i: 1 \dots R$) del primer conjunto al que se puede denotar como I o conjunto de individuos o sujetos clasifica u otorga una puntuación s_{ij} a cada uno de los elementos j (con $j: 1 \dots N$) del segundo conjunto denominado J o conjunto de objetos o estímulos. El modelo Unfolding supone que los sujetos perciben los objetos de la misma manera pero que difieren en lo que consideran la combinación ideal de los atributos de los mismos. Proporciona una representación conjunta de los individuos y los estímulos en el mismo espacio euclídeo de dimensión M , en la que la distancia desde el punto que representa a un individuo i (también llamado punto ideal) hasta el punto o lugar que representa a un estímulo j está inversamente relacionada con la correspondiente puntuación. Así, un valor s_{ij} que indique una gran preferencia del individuo i hacia el estímulo j , estará asociado a una distancia pequeña entre los puntos que representan al individuo i y el estímulo j , y lo mismo sucede al contrario: valores s_{ij} que indiquen poca preferencia se asociarán con distancias grandes.

Una aproximación inicial: Modelo de Punto-Ideal e Isopreferencias

El ejemplo clásico de Green y Rao (1972) permite explicar las nociones básicas del **modelo Unfolding (también llamado modelo del “punto ideal”)** y las **isopreferencias**.

En este experimento, se pidió a 21 individuos y a sus parejas (42 sujetos en total) que ordenasen de mayor a menor preferencia 15 productos para el desayuno. Una vez recopilados los datos de todos los individuos, se construyó una matriz de dimensión 42x15, cuyos elementos por fila eran las puntuaciones otorgadas a cada producto desde 1 (más preferido) a 15 (menos preferido). Estas puntuaciones pueden expresar de alguna manera cierta noción de proximidad, es decir, la cercanía de cada producto a un producto de desayuno óptimo.

Si se retrocede de nuevo al MDS, se puede recordar que la matriz de similitudes/disimilitudes en MDS clásico es una matriz cuadrada, donde las entidades representadas en filas y columnas son las mismas. En el caso del Unfolding, la matriz de entrada es una matriz rectangular de preferencias en dos vías y dos modos (sujetos y estímulos). Pero se puede conseguir que la matriz $R \times N$ se convierta en una matriz cuadrada $(R+N) \times (R+N)$, simplemente yuxtaponiendo sujetos y estímulos uno a continuación del otro tanto en filas y columnas. De esta manera, la matriz estaría formada por cuatro bloques, dos con las puntuaciones de preferencia dadas por los individuos para el conjunto de estímulos y otros dos con datos ausentes (ya que no existen puntuaciones entre los elementos de un mismo grupo). Así, en esta matriz (**Fig. 3**) sólo se recogen datos de proximidad entre ambos conjuntos (individuos y objetos), pero no dentro del mismo conjunto.

	Estímulos	Sujetos
Estímulos	Datos ausentes	
Sujetos		Datos ausentes

Fig. 3. Matriz de preferencias

Con los datos de partida y aplicando el algoritmo SMACOF de minimización del Stress, la solución del modelo Unfolding vendría dada por una configuración conjunta de $(R+N)$ puntos que representan a R individuos y N objetos. El modelo Unfolding también se conoce como **Modelo del Punto Ideal**, pues éste asume que todos los individuos comparten el mismo espacio perceptual para la elección de los objetos (es decir, los individuos comparten el mismo espacio psicológico para la elección de estímulos). La proximidad o cercanía entre un punto que representa a un objeto y un punto que representa a un individuo ("**punto ideal**" o **imagen**) refleja la preferencia que siente el individuo por dicho objeto y las diferencias entre individuos se adivinan a través de las distancias entre los "puntos ideales" de la representación. Los contornos de **isopreferencias** (según notación de Borg) son los círculos concéntricos centrados en un punto ideal y engloban al conjunto de objetos o estímulos por los que el individuo representado por ese punto ideal muestra una preferencia similar.

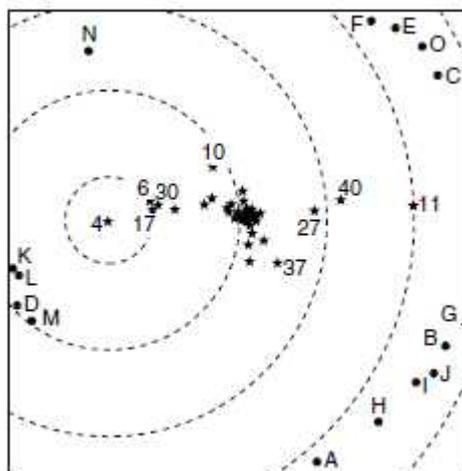


Fig. 4. Un ejemplo de Solución Unfolding (Green y Rao, 1972)

Soluciones Degeneradas o Triviales.

A priori el modelo del punto ideal parece solucionar la cuestión planteada en este ejemplo, se encuentra una configuración que ajusta de manera adecuada los datos de elección, pero ¿hasta qué punto es esto correcto y es válido en cualquier conjunto de datos?

Para responder a esta pregunta, será necesario tener en cuenta una serie de aspectos y consideraciones propias de MDS.

El MDS utiliza datos de proximidad para reproducir la configuración de puntos pudiendo distinguirse (como ya se vio en el apartado de MDS) entre datos

incondicionales y condicionales, lo que resulta también aplicable al Unfolding. Con datos incondicionales, se permite que cualquier dato de proximidad se pueda comparar con otro dato de proximidad dentro de la matriz de datos, pero para datos condicionales esto no es posible. La matriz de preferencias que se muestra en la (**Fig. 3**), está formada por filas cuyos elementos son puntuaciones correspondientes a individuos diferentes y se ha de tener en cuenta que:

- 1) El orden de los estímulos puede estar influenciado por aspectos psicológicos diferentes.
- 2) Las comparaciones entre individuos se muestran complicadas dada la estructura de los datos.
- 3) Si se tiene en cuenta un Unfolding Incondicional, todas las puntuaciones 1 deberán ser transformadas en distancias más pequeñas que aquellas que representan las puntuaciones 2 y así sucesivamente, para todas las filas de la matriz, mientras que en el caso del Unfolding Condicional se requiere sólo que una puntuación 1 se transforme en una distancia más pequeña que la distancia que representa a la puntuación 2 dentro de la misma fila de manera independiente en cada fila de la matriz.

La figura obtenida a través de un Unfolding Condicional por filas (**Fig. 4**), parece que transforma de manera adecuada el orden de elección en proximidades, pero para ciertos individuos muchos productos se encuentran dentro de una de las regiones de isopreferencia y esto puede ser síntoma de **soluciones degeneradas**, porque de alguna manera no se ve reflejada correctamente su elección.

Las soluciones degeneradas, que aparecen también en MDS clásico, son provocadas por inconsistencias en los datos o por una incapacidad del programa de análisis informático para alcanzar una solución adecuada, ya que, en ocasiones, se alcanza un valor muy pequeño del Stress con independencia de la relación que exista entre las proximidades de partida y las distancias de los puntos de la configuración obtenida. Las soluciones degeneradas se pueden corregir imponiendo restricciones adicionales en la función de pérdida, por ejemplo, aplicando una nueva medida derivada del Stress. Más adelante se profundizará en este asunto.

Además de las soluciones degeneradas que son propias del MDS y que el Unfolding también las hereda, aparecen otro tipo de soluciones inesperadas llamadas **soluciones triviales** que son independientes del tipo de Unfolding aplicado y del tipo de datos a analizar.

En este punto es conveniente comentar que puede haber cierta controversia acerca de la distinción entre solución degenerada y trivial, ya que puede que no se encuentre distinción entre ambos conceptos. Borg (2005) es bastante ambiguo respecto de la

diferencia, sin embargo en Busing (2010) el autor comenta que DeLeeuw (1983) hace una distinción clara entre ambas soluciones:

- i. Una solución trivial tiene $\text{Stress}=0$ y proporciona una solución no interpretable dentro de la configuración conjunta, pero se puede evitar con una normalización adecuada.
- ii. Una solución degenerada, sin embargo, a menudo no tiene por qué alcanzar un Stress cero, tampoco proporciona soluciones interpretables pero no es posible evitarla solo con una normalización.

En algunas ocasiones, las soluciones triviales y degeneradas pueden presentarse en forma de solución trivial equidistante: los puntos de uno o ambos conjuntos (individuos y objetos) están dispuestos en dos agrupaciones compactas, de forma que las distancias entre dos puntos cualesquiera es siempre la misma, como se puede apreciar a continuación.



Fig. 5. Un ejemplo de solución trivial equidistante

Para evitar este tipo de soluciones triviales equidistantes, Kruskal (1968) y Kruskal y Carroll (1969) propusieron una variante del Stress, llamada Stress-2:

$$\sigma_2(\mathbf{X}) = \frac{\sum_{i < j} w_{ij} [d_{ij}(\mathbf{X}) - \bar{d}_{ij}]^2}{\sum_{i < j} w_{ij} [d_{ij}(\mathbf{X}) - \bar{d}]^2} \quad (6)$$

donde \bar{d} es la media de todas las distancias $d_{ij}(\mathbf{X})$, con $i < j$. Técnicamente la minimización de esta nueva medida resuelve las soluciones triviales equidistantes (ya que para este tipo de soluciones el denominador tiende a cero y, por tanto, el Stress-2 es grande), pero no resuelve otro tipo de soluciones triviales que se dan en el Unfolding Condicional por filas, propio de datos de preferencia.

En el Unfolding Condicional por filas, las transformaciones que se utilizan son ordinales y los datos se transforman independientemente fila a fila. En este caso, la minimización del Stress puede conducir a soluciones en las que todos los estímulos se representan mediante un único punto y los individuos (filas de la matriz) se representan en la configuración mediante puntos diseminados por todo el espacio. La razón por la cual se puede obtener una solución de este tipo reside en el hecho de que la transformación condicional por filas permite diferentes distancias entre los puntos que representan a las filas y el punto que representa a las columnas mientras que mantiene iguales las distancias dentro de las filas. El numerador del Stress-2 es cero ya que $d_{ij}(X) = \hat{d}_{ij}$ dentro de cada fila de la matriz, pero el denominador no, ya que las distancias de una fila al punto único que representa a los estímulos difieren de fila a fila.

Las soluciones triviales y los enfoques propuestos para evitar su aparición se discuten más profundamente en el siguiente capítulo.

Como se comentó en párrafos anteriores, las soluciones degeneradas son propias del MDS y además también aparecen en el Unfolding debido a problemas de minimización en la función de pérdida o inconsistencias en los datos.

Un ejemplo típico se muestra en (Fig. 6), en la que los puntos del conjunto I (filas de la matriz) son transformados en puntos que yacen unos superpuestos sobre otros, mientras que los puntos de J (columnas) son transformados en puntos que yacen alrededor de una curva. Esta configuración es propia de un Unfolding ordinal incondicional, cuya solución se ha obtenido minimizando el Stress-1.

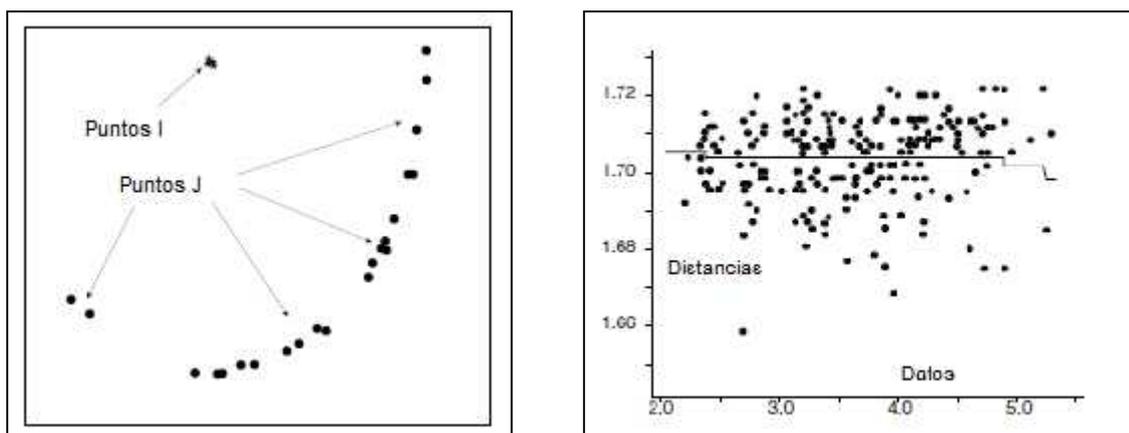


Fig. 6. Un ejemplo de solución degenerada para Unfolding ordinal incondicional y diagrama de Shepard asociado con el Stress-1.

El diagrama de Shepard muestra que los puntos que representan las distancias se dispersan muy suavemente respecto de la línea central (línea de ajuste), por lo que la suma de los cuadrados de la distancia de cada punto respecto de la línea (numerador del Stress-1) es mucho más pequeño que la suma de los cuadrados de las distancias de todos los puntos (denominador del Stress-1).

Si se intenta utilizar el Stress-2 propuesto por Kruskal (6) sobre el mismo conjunto de datos (Fig. 7), la representación que se obtiene no mejora la calidad en el sentido que de nuevo se obtiene una solución degenerada.

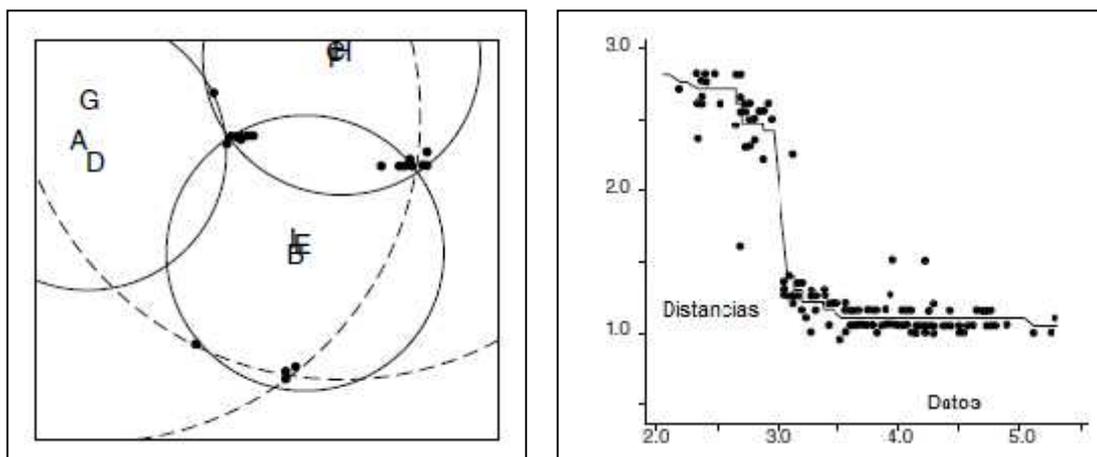


Fig. 7. Un ejemplo de solución degenerada para Unfolding ordinal incondicional y diagrama de Shepard asociado con el Stress-2.

El diagrama de Shepard asociado a la configuración obtenida muestra que existe una “degeneración” de las distancias, puesto que las puntuaciones menores de 3 se transforman en distancias de aproximadamente 2,5 mientras que las puntuaciones mayores de 3 lo hacen en distancias de alrededor de 1,2 y todos los puntos se agrupan cerca de la línea de ajuste, lo cual indica un numerador del Stress- 2 muy pequeño o próximo a cero.

En la figura de la izquierda, que representa la configuración obtenida, se observa que los estímulos (etiquetados con letras) se agrupan en tres clusters mientras que los individuos (etiquetados mediante puntos) también se presentan agrupados. Se han dibujado tres círculos en trazo continuo con centro los estímulos de cada cluster y con radio la distancia al grupo de individuos más cercano, y otros dos círculos en trazo discontinuo con centro en AGD y CFH, respectivamente y con radio la distancia al grupo de individuos más alejado. Se puede observar que tanto los círculos continuos como los discontinuos tienen el mismo radio, lo que demuestra que la transformación

aplicada a los datos calcula dos distancias tal y como muestra el diagrama de Shepard. Borg y Groenen afirman, además, que si la matriz se transpone en el sentido de que los estímulos aparecen en filas y los individuos en columnas, la solución sería la misma (dos distancias) intercambiándose en la configuración individuos (centro de los clusters) y estímulos (sobre los círculos continuos).

Por lo tanto, ya que sobre un Unfolding Incondicional Ordinal la minimización del Stress-1 o el Stress-2, no resuelve el problema de las soluciones degeneradas, Borg y Groenen plantean utilizar un Unfolding métrico con datos de preferencia supuestos en una escala de intervalo.

Sin embargo, al aplicar de nuevo un Unfolding Incondicional de intervalo, la configuración de puntos que se obtiene minimizando el Stress-1 es parecida a la de la (Fig. 6), con la salvedad de que el diagrama de Shepard mostraría una recta de regresión con pendiente positiva, lo contrario a lo esperado (pendiente negativa) dado que la recta debería decrecer para preservar la interpretación de que los individuos que puntúan más alto deberían estar más próximos en distancia. De todos modos esta solución tampoco parece satisfacer las expectativas, ya que de nuevo lleva a una solución degenerada que no se corresponde con los datos originales.

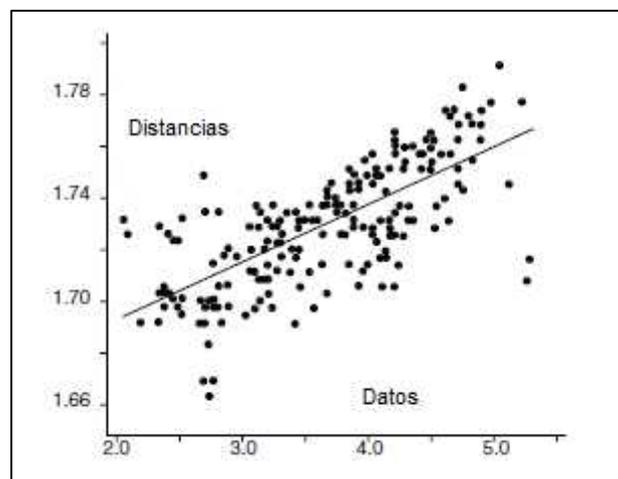


Fig. 8. Diagrama de Shepard para Unfolding Incondicional de intervalo con Stress-1.

Aplicando una minimización del Stress-2 en el mismo Unfolding Incondicional de intervalo anterior, a priori, tanto la configuración de puntos obtenida como el diagrama de Shepard, no parecen mostrar soluciones degeneradas, los puntos parecen ajustarse muy bien a la línea de regresión y no existen discontinuidades. Sin embargo al comprobar la “realidad” de la solución, la recta de regresión tiene una pendiente negativa (contrario a lo esperado) y la configuración de los datos, parece ser que no

representa de manera fiel las preferencias de cada individuo, es decir, estímulos más preferidos no son representados junto a los individuos que los han elegido.

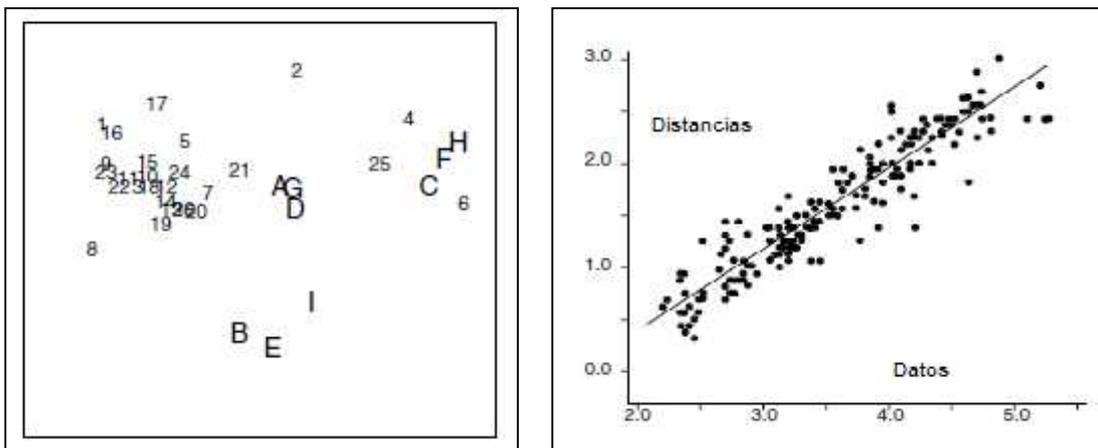


Fig. 9. Solución Unfolding Incondicional de intervalo y diagrama de Shepard con Stress-2.

Por lo tanto, la utilización del Stress-2 como medida de ajuste para evitar las soluciones degeneradas no es del todo correcta, ya que en el caso del Unfolding Ordinal no se consiguen eliminar y si se utiliza una aproximación métrica (por ejemplo, a través de un Unfolding de intervalo) aunque tiende a desaparecer, se obtiene una configuración que no se corresponde con los datos iniciales.

Configuración de Puntos en el Espacio Común (Regiones isótonas e indeterminaciones)

Otro punto muy importante que se debe tener en cuenta a la hora de realizar un análisis Unfolding Condicional de datos ordinales es la calidad de representación de los datos. Hay que hacer especial hincapié en la importancia que tiene la posición de los puntos de un conjunto con respecto al resto de puntos de su mismo conjunto y respecto a los puntos del segundo. Según Borg y Groenen, la mejor distribución relativa de puntos y objetos es aquella en la que ambos conjuntos están distribuidos alrededor del espacio, lo que implica que hay individuos con diferentes alternativas de elección (preferencias) y cada uno de ellos tiene un objeto más preferido entre éstas.

La existencia de una configuración en un espacio de dos dimensiones, en la que individuos y objetos se distribuyan correctamente conservando la ordenación de preferencias, fue probada por Coombs (1964), que demostró que si existen n objetos en un espacio MDS de dimensión $n-1$, entonces todas las regiones isótonas de los puntos ideales son abiertas.

En el caso del Unfolding Multidimensional, el espacio conjunto de dimensión p puede ser dividido en regiones isótonas por los hiperplanos definidos por cada par de

estímulos (**Fig. 10**) y de acuerdo al orden de preferencia, todos los individuos que yacen en una determinada región isótónica, tienen la misma preferencia por el conjunto de estímulos, por ejemplo en (**Fig. 10**) los individuos que tengan el mismo orden de preferencia $D > C > A > B$, serán representados en el área subrayada.

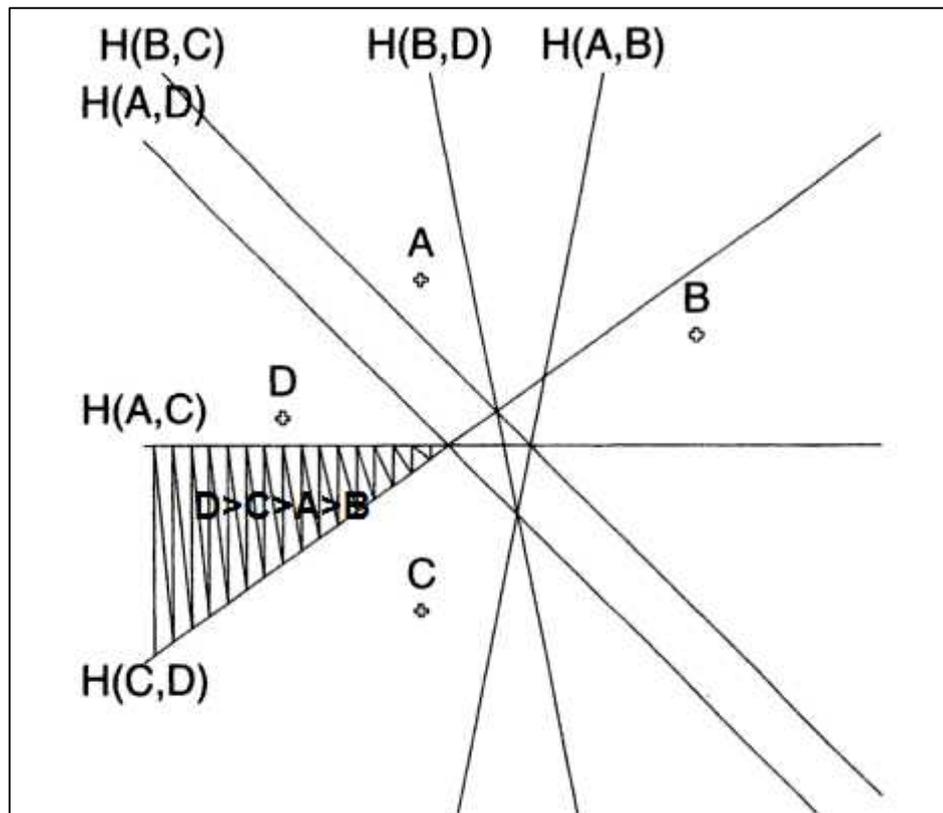


Fig. 10. Ejemplo de Regiones Isótónicas con 4 estímulos en un espacio de 2 dimensiones (Cox, 2000)

Conforme aumenta el número de estímulos, el número de hiperplanos crece y el área descrita por las regiones isótónicas se reduce, reduciendo también la incertidumbre de la representación de las coordenadas de los individuos ya que estarían forzados por las preferencias y las regiones isótónicas que les pertenecen.

En ese sentido cuanto más pequeña se haga el área de la región isótónica más se reduce la incertidumbre en la representación de los individuos respecto de sus preferencias y, por tanto, en la unicidad de la solución obtenida.

Otras consideraciones en Unfolding Multidimensional

Hasta ahora se han tratado dos de los puntos más importantes que cualquier investigador de Unfolding debería conocer: las soluciones triviales y las indeterminaciones en la representación conjunta de puntos en el espacio. No obstante, se podrían añadir ciertos aspectos de índole psicológica, tan importantes como las

anteriores, si se quiere obtener una configuración de calidad que reproduzca lo más fielmente posible los datos de entrada.

En primer lugar, hay que tener en cuenta las dimensiones que subyacen en la interpretación de los resultados o en la elección de los estímulos por parte de los individuos para no llegar a conclusiones erróneas. Se aconseja tener especial precaución en no colapsar el espacio conjunto debido a la existencia de dimensiones correlacionadas, es decir evitar realizar la elección de estímulos en dimensiones que impliquen psicológicamente lo mismo y hacer una minuciosa selección de estímulos para realizar el juego de preferencias.

En segundo lugar, el Unfolding Multidimensional asume que en la elección de los estímulos, todos los individuos perciben el juego de la misma manera, es decir, la colección de objetos significa psicológicamente lo mismo para todos, la configuración de los estímulos es única y las diferencias entre individuos se reducen únicamente a las distancias entre los puntos ideales, no existen dimensiones psicológicas añadidas que tengan interpretación en la configuración. Si esto no se tiene en cuenta, la interpretación del resultado llevará a conclusiones erróneas.

Soluciones Triviales, ¿cómo resolverlas?: Tres enfoques distintos

Tal y como se ha comentado en el apartado anterior, el principal problema que surge a la hora de obtener una representación correcta y válida de la configuración de los puntos de ambos conjuntos en el espacio, es el de las soluciones triviales.

Una de las alternativas propuestas para intentar solucionar este problema es el Stress-2 que, como se ha visto anteriormente, no las evita del todo. Borg y Groenen proponen y discuten tres tipos de enfoques para tratar de resolver este tipo de soluciones:

- ✓ Ajuste o adaptación de los datos
- ✓ Ajuste de la transformación
- ✓ Modificación de la función de pérdida

Ajuste o adaptación de los datos

El primer enfoque para evitar una solución trivial consiste en asegurar que la función de transformación no pueda tener un término independiente distinto de cero y una pendiente nula, para lo cual se adaptan los datos de la matriz de entrada, a través de dos alternativas:

- **Aproximación Ordinal -Razón (*Ordinal-Ratio Approach*)**

La idea de esta alternativa es utilizar una transformación ordinal ($\delta_{ij} < \delta_{rs} \rightarrow f(\delta_{ij}) \leq f(\delta_{rs})$) y una de razón ($f(\delta_{ij}) = b\delta_{ij}$) simultáneamente sobre los mismos datos y aproximar las disparidades obtenidas con ambas transformaciones mediante una matriz de distancias única.

Se puede obtener la función de pérdida para cada transformación, $L(o)$ para la transformación ordinal y $L(r)$ para la de razón, lo que permite definir una función de pérdida conjunta como:

$$L = aL(o) + bL(r) \quad (7)$$

con $a, b > 0$ y $a + b = 1$ y las siguientes consideraciones:

- L es cero si y sólo si $L(o)$ y $L(r)$ lo son a la vez.
- Si L tiende a cero, entonces o $L(o)$ o $L(r)$ o ambos tienden a cero o bien uno de ellos lo hace y el otro no alcanza un valor grande.
- La transformación ordinal trata siempre de conservar el orden y la transformación de razón no tiene término independiente, así que la transformación trivial (constante) no se puede dar.

La desventaja de esta tentativa de resolución es que, quizás debido a la transformación de razón, sería más ventajoso utilizar como datos de partida disimilaridades en vez de similaridades. Cuando se trabaja con similaridades, se requiere que la transformación de razón tenga pendiente negativa, de modo que haga corresponder distancias pequeñas a similaridades grandes. Esto puede dar lugar a disparidades negativas, que no pueden aproximarse adecuadamente a distancias no negativas. Por ello, al trabajar con similaridades se puede optar por convertirlas en disimilaridades antes de realizar el Unfolding, o bien por reemplazar la transformación de razón por una transformación de intervalo ($f(\delta_{ij}) = a + b\delta_{ij}$) y considerar la función de pérdida $L = aL(o) + bL(i)$, con $L(i)$ la función de pérdida para la transformación de intervalo.

Tal y como comentan Borg y Groenen, no está garantizado que la aproximación ordinal-intervalo funcione siempre (es decir, evite las soluciones triviales), aunque estos autores aplican con éxito esta aproximación a un conjunto de datos.

- **Ampliación de la matriz de datos con proximidades entre entidades de la misma clase (*Augmenting the Within-Objects Blocks*)**

Esta alternativa para adaptar los datos iniciales se basa en la idea de que el Unfolding es equivalente a un MDS con una matriz de proximidades incompleta (**Fig. 3**) y sugiere ampliar esta matriz calculando los datos de uno (o ambos) bloques con datos faltantes. Steverink, Van der Kloot y Heiser (2002) propusieron reemplazar los datos ausentes del bloque de los sujetos mediante la distancia de Kemeny, ya que ésta parece especialmente adecuada para el Unfolding con datos de preferencia.

La distancia de Kemeny (Kemeny, 1959) se calcula de la siguiente forma:

Para un individuo i y un par de estímulos k, l cualesquiera, se define la función:

$$z_i(k, l) = \begin{cases} 1 & \text{si } z_i(k) > z_i(l) \\ 0 & \text{si } z_i(k) = z_i(l) \\ -1 & \text{si } z_i(k) < z_i(l) \end{cases}$$

Entonces la distancia de Kemeny entre los individuos i y j se calcula como:

$$d_{ken}(i, j) = \sum_{k < l} |z_i(k, l) - z_j(k, l)| \quad (8)$$

En principio, una transformación ordinal sobre la matriz con reemplazo de datos ausentes por los valores correspondientes de la distancia de Kemeny elude la solución trivial equidistante, pero no ofrece una representación válida conjunta de individuos y estímulos.

Otro punto importante a tener en cuenta es que si la transformación sobre la matriz aumentada se hace condicionando por filas, la representación conjunta de ambos da una buena calidad de representación de las distancias entre los elementos de cada grupo, pero no de las distancias entre elementos de distinto grupo (es decir, de las preferencias de los individuos por los estímulos). Para dar solución a este último punto, Borg y Groenen (2005) proponen reemplazar también el segundo bloque de datos ausentes pero esta vez calculando la distancia de Manhattan⁵ a partir de cada vector de preferencias. Además llevan a cabo dos pasos adicionales: primero, transforman las distancias de ambos bloques en euclídeas mediante la raíz cuadrada y, segundo, hacen que el rango de los valores en los bloques reemplazados sea igual

⁵La **distancia city block** o **Manhattan** entre dos puntos es la suma de las diferencias absolutas de sus coordenadas : $D(X, Y) = \sum_{i=1}^N |x_i - y_i|$

dividiendo todos los elementos de cada bloque por su máximo valor, para asegurar que ambos bloques son equivalentes.

Con esta aproximación más los pasos adicionales anteriormente descritos, Borg y Groenen consiguen una calidad aceptable en la representación de ambos conjuntos, aunque muestran cierta cautela con esta metodología e invitan a futuros investigadores a mejorar este tipo de aproximación.

Ajuste de la transformación (*Adjusting the Transformation*)

El segundo enfoque para evitar las soluciones triviales consiste en restringir las transformaciones posibles a aquellas que no tengan simultáneamente un término independiente no nulo y una pendiente nula.

La transformación de razón $f(\delta_{ij}) = b\delta_{ij}$ satisface esta restricción, pero no es del todo eficiente ya que obvia la naturaleza ordinal de los datos originales usados frecuentemente en Unfolding.

Kim, Rangaswamy y DeSarbo (1999) proponen una idea parecida a la anterior, desarrollada en dos etapas. La primera etapa consiste en transformar los datos originales mediante una transformación λ_{ij} que cumpla las siguientes propiedades:

- i. λ_{ij} debe ser estrictamente monótona.
- ii. En cada fila, el estímulo más preferido se transforma en cero.
- iii. La transformación λ_{ij} es la misma para todas las filas.

La segunda etapa consiste en la utilización de los datos transformados para el cálculo del Stress.

Esta transformación funciona debido a que los λ_{ij} de cada columna no pueden ser iguales; así mismo, los autores disertan acerca de la elección de dicha transformación λ_{ij} , ya que cualquier función que cumpla las tres propiedades anteriormente mencionadas puede ser candidata aunque, dependiendo de cuál sea la función elegida, las soluciones obtenidas sobre los mismos datos serán diferentes. Además opinan que sería conveniente encontrar otro tipo de funciones distintas de la transformación de razón que funcionen bien para preferencias expresadas en escala ordinal. Sugieren también la aplicación de una regresión monótona suavizada⁶ propuesta por Heiser (1985,1989) como una posible medida que cumple los requisitos

⁶La **regresión suavizada** es un modelo de regresión no paramétrico en el que la función de regresión que relaciona ciertas variables no tiene una forma predeterminada o conocida y sólo se puede suponer sobre ella ciertas hipótesis de suavidad en el sentido de continuidad y diferenciabilidad.

establecidos y además evita soluciones triviales en la representación conjunta de individuos y estímulos.

Ajustes en la Función de Pérdida (*Adjustments to the Loss Function*)

El tercer enfoque recogido por Borg y Groenen se centra en la función de pérdida⁷ evaluada a través del Stress.

En primer lugar tal y como se vio en el apartado anterior, se puede utilizar el Stress-2 (6) para evitar soluciones triviales, pero en ocasiones, si se utiliza una transformación lineal sobre los datos no se podría garantizar que éstas no ocurrieran. Por ese motivo, los autores recogen varias propuestas relacionadas con la función de pérdida para evitar su aparición.

- Asignar pesos (*Weighting Strategies*)

La estrategia ideada por DeSarbo y Rao (1984) consiste en evitar la ocurrencia de soluciones triviales ponderando los datos iniciales en el cálculo del Stress bruto. Dicha ponderación tendría la expresión: $w_{ij} = \delta_{ij}^{-p}$, donde δ_{ij} son las puntuaciones de preferencia y el exponente $p > 0$ determina la influencia de un determinado objeto, de manera que los estímulos más preferidos tienen pesos pequeños y los menos preferidos, pesos grandes. Si se observa la (Fig. 11) las curvas correspondientes a un exponente p : 1,..5, informan acerca de la influencia de ciertos objetos; así, por ejemplo, si $p=2$ el residuo del objeto preferido en segundo lugar solo contribuye con un peso del 25% en la función de Stress y para $p=5$ sólo con un 3%. Esto significa que únicamente los estímulos más preferidos contribuyen al Stress y por tanto existen pocas restricciones efectivas entre los puntos que representan a objetos e individuos con la consiguiente indeterminación de la solución. La calidad de la solución depende principalmente de la calidad de la configuración inicial.

Además al no haber restricciones sobre la transformación (por ejemplo, puede usarse una transformación constante), las diferentes puntuaciones de preferencia pueden ser transformadas en distancias iguales y, por lo tanto, no se evitan las soluciones triviales

⁷ Precisa la **pérdida** en la que se incurre al hacer la representación gráfica en un espacio de dimensión reducida.

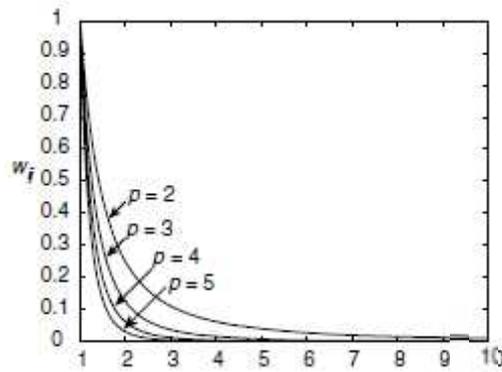


Fig. 11. Representación gráfica de la función de pesos w_{ij} (Borg y Groenen, 2005)

- Penalización del término independiente (*Penalizing the Intercept*)

Otro ajuste sobre la función de pérdida para eludir soluciones triviales la plantea Busing (2005), aunque realmente sólo se ha probado que es efectiva en transformaciones de tipo lineal $(a + b\delta_{ij})$. Su idea consiste en añadir una penalización en la función de pérdida, en el caso de que la transformación anterior tuviera un término independiente de gran valor.

La nueva función de pérdida tendría la siguiente expresión:

$$l(a, b, X) = \sum_{i,j} [a + b\delta_{ij} - d_{ij}(X)]^2 + \omega a^2 \quad (9)$$

siendo $\omega > 0$ el factor que determina la magnitud de la penalización usada. Si $\omega = 0$ la expresión anterior se reduce a la ya conocida del Stress y en el caso extremo de que $\omega \rightarrow \infty$, obliga a que el coeficiente a sea cero, siendo la transformación en este caso de tipo razón $(b\delta_{ij})$.

Para la aplicación de la penalización sobre el término independiente es necesario trabajar, además de con transformaciones lineales, con disimilaridades. Si se trabajase con similaridades cabría esperar una transformación con un término independiente grande y una pendiente negativa, pues similaridades grandes se transforman en distancias y disparidades pequeñas mientras que similaridades pequeñas se ven transformadas en distancias y disparidades grandes. Esto conlleva que la transformación arrastre un término independiente de gran valor lo que contradiría la idea de penalizar el término independiente para que no alcance gran valor.

De todos modos, Borg y Groenen (2005) consideran que apenas ninguna transformación puede ser utilizada en esta aproximación con buenos resultados, excepto las funciones lineales o de tipo spline⁸.

Sin embargo, al aplicar esta aproximación a un conjunto de datos usando un Unfolding Lineal, concluyen que, al menos bajo el supuesto de transformación lineal, la penalización del término independiente parece efectiva para evitar la solución trivial.

- **PREFSCAL: Penalización de disparidades iguales (*Penalizing Equal d-hats*)**

La tercera y última aproximación para eludir soluciones triviales dentro del grupo de ajustes en la función de pérdida, debida a Busing et al. (2005), usa el modelo PREFSCAL (PREFerence SCALing) que está implementado en el programa SPSS para realizar Unfolding Multidimensional.

La idea fundamental de la que parten Busing y colaboradores, es que las soluciones triviales están caracterizadas por disparidades constantes y que, por lo tanto, una manera de evitarlas sería penalizar la función de Stress para \hat{d}_{ij} iguales. La ventaja de este método es que es válido para cualquier transformación de datos y la configuración resultante puede ser interpretada a través del modelo del punto ideal.

En primer lugar, Busing et al. proponen identificar las disparidades iguales, a través del coeficiente de variación de Pearson:

$$v(\hat{d}) = \frac{\text{desv. estándar}(\hat{d})}{\text{media}(\hat{d})} = \frac{(\sum_k (\hat{d}_k - \bar{\hat{d}})^2)^{-1/2}}{K^{-1} \sum_k \hat{d}_k} \quad (10)$$

donde $\bar{\hat{d}} = K^{-1} \sum_k \hat{d}_k$ y k recorre todas las disparidades \hat{d}_{ij} .

Este coeficiente de variación es una medida que indica cuánto varían las disparidades respecto de la media; además el coeficiente de variación es invariante con respecto a la escala, por tanto, $v(a\hat{d}) = v(\hat{d})$ dado $a > 0$.

En segundo lugar, Busing et al. proponen la utilización del coeficiente de variación como penalización del Stress. Su modelo PREFSCAL minimiza una variación de la función de Stress, que se llama Stress penalizado y se define como:

⁸Una **función spline** está formada por varios polinomios, cada uno definido sobre un subintervalo, que se unen entre sí obedeciendo a ciertas condiciones de continuidad.

$$p(\bar{d}, X) = \sigma n^\lambda(\bar{d}, X) \left(1 + \frac{\omega}{v^2(\bar{d})} \right) \quad (11)$$

donde $n(\bar{d}, X)$ es el Stress normalizado⁹, y ω y v son los parámetros de penalización que deben especificarse bajo las restricciones $0 < \omega < 1$ y $v > 0$. Cuanto más próximo sea ω a cero más peso tiene $\left(1 + \frac{\omega}{v^2(\bar{d})} \right)$. El parámetro v , a su vez, decidirá la influencia del término anterior en el Stress penalizado, ya que cuanto mayor valor alcance v , mayor peso tendrá $\left(1 + \frac{\omega}{v^2(\bar{d})} \right)$ en la medida global. Busing et al. recomiendan elegir $\omega = 0,5$, aunque también sugieren que ω podría modificarse dependiendo de los datos. Cuando el coeficiente de variación está próximo a cero está claro que la expresión $\left(1 + \frac{\omega}{v^2(\bar{d})} \right)$ toma un valor grande y por lo tanto, al minimizar $p(\bar{d}, X)$, el algoritmo no permitirá que se obtengan \bar{d}_{ij} constantes.

Una de las ventajas que Busing et al. destacan sobre la utilización del Stress penalizado es que cuanto más alejado se esté de la solución trivial, menos influencia tiene el término de penalización y más influencia tiene $n^\lambda(\bar{d}, X)$. Además el Stress penalizado vale cero cuando el Stress normalizado es cero, es decir, en el caso de soluciones no triviales; de ese modo, si el problema de Unfolding tiene solución y ésta es no trivial, el Stress penalizado la encuentra.

Además como se comentó anteriormente al ser $n(\bar{d})$ invariante por escala, también lo es el Stress Penalizado y no afecta a la escala en la que se expresen X y \bar{d} . Sin esa propiedad (invariancia de escala), el Stress Penalizado sería sensible al tamaño (n^0 de individuos y estímulos) y por ende al tamaño de la matriz. Por lo tanto, el algoritmo PREFSCAL es independiente del número de filas (individuos) y columnas (estímulos) de la matriz así como de la normalización de \bar{d} .

En el caso de transformaciones condicionales por fila, para evitar \bar{d}_{ij} constantes en cada fila, Busing et al. indican que el término de penalización debe ser más grande por lo que proponen un nuevo Stress penalizado:

$$p_{rc}(\bar{d}, X) = \sigma n^\lambda(\bar{d}, X) n^{2-1} \sum_{i=1}^{n^2} \left(1 + \frac{\omega_i}{v^2(\bar{d}_i)} \right) \quad (12)$$

⁹Stress Normalizado $n(\bar{d}, X) = \frac{\sum_{i,j} w_{ij} |d_{ij}(X) - \delta_{ij}|^2}{\sum_{i,j} w_{ij} \delta_{ij}^2}$

donde n_2 es el número de filas de la matriz y \hat{d}_i contiene las disparidades de la fila i . Como en el caso anterior, si $v(\hat{d}_i)$ es cercano a cero, $-2(\hat{d}_i)$ toma valores grandes y el término $n_2^{-1} \sum_{i=1}^{n_2} \left(1 + \frac{v(\hat{d}_i)}{v^2(\hat{d}_i)}\right)$ también es grande.

Al aplicar PREFSCAL, y el correspondiente coeficiente de penalización mediante una transformación incondicional (Fig. 12) o condicional (Fig. 13), en ambos casos la configuración de puntos parece ser correcta, sin soluciones triviales, los diagramas de Shepard asociados tienen pendiente negativa y la nube de puntos parece ajustarse bien a la línea central. Por lo tanto se puede deducir que PREFSCAL y la penalización usando el coeficiente de variación propuesta por Busing et al. funcionan razonablemente bien, tanto para Unfolding Incondicional (Fig. 12) como para Unfolding Condicional (Fig. 13).

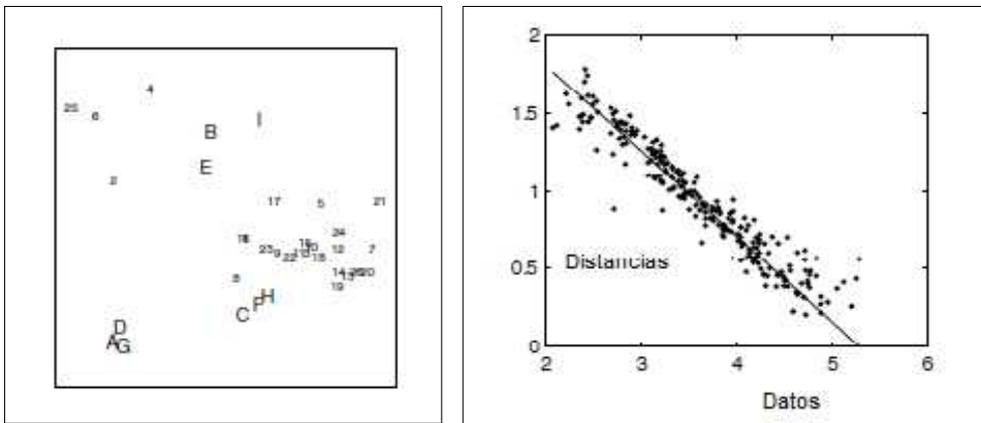


Fig. 12. Ejemplo de Solución Unfolding Lineal mediante PREFSCAL y diagrama de Shepard

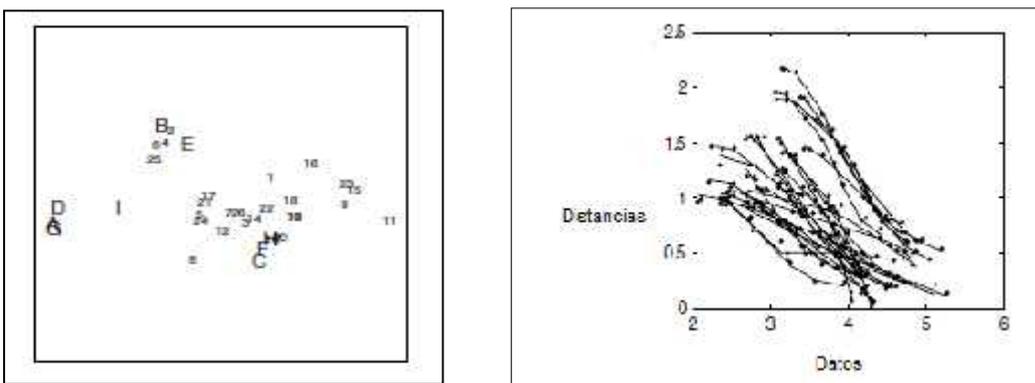


Fig. 13. Ejemplo de Solución Unfolding Condicional mediante PREFSCAL y diagrama de Shepard

Modelos de Unfolding

Siguiendo la clasificación realizada por Zarboski (2010) y DeSarbo y Rao (1984), se pueden diferenciar dos grupos de modelos de Unfolding: los del **Unfolding Interno** y los del **Unfolding Externo**. Ambos conceptos se deben a Carroll (1972).

- **Unfolding Interno:** En estos modelos, las configuraciones X de los estímulos e Y de los sujetos (puntos ideales) se obtienen a partir de una matriz de preferencias tal y como se describe en (Fig. 3) minimizando la función de Stress dada por:

$$Stress = \sqrt{\frac{\sum_{i < j} [d_{ij} - \hat{d}_{ij}]^2}{\sum_{i < j} d_{ij}^2}} \quad (13)$$

donde:

- $d_{ij} = \sqrt{\sum_{a=1}^R (x_{ja} - y_{ia})^2}$ son las distancias entre los individuos de coordenadas y_{ia} y los estímulos de coordenadas x_{ja} .
- $\hat{d}_{ij} = f(\delta_{ij})$ es la disparidad obtenida mediante una regresión monótona de d_{ij} sobre δ_{ij} .

Cuando las preferencias están expresadas en escala ordinal, las disparidades \hat{d}_{ik} han de cumplir la siguiente restricción monótona:

$$\delta_{ij} \leq \delta_{i'j'} \rightarrow \hat{d}_{ij} \leq \hat{d}_{i'j'} \quad (14)$$

La solución se obtiene a través del algoritmo de mayorización **SMACOF**, que actualiza las coordenadas de estímulos y sujetos en cada paso de forma que se reduzca progresivamente el Stress (Heiser, 1981).

SMACOF (Scaling by Majorizing a Complex Function, DeLeeuw y Heiser (1977))

El algoritmo de mayorización SMACOF reduce el Stress (13) mediante sucesivas iteraciones. Tras K iteraciones se actualizan X e Y (matrices con las coordenadas de estímulos e individuos en el espacio conjunto) de la siguiente forma:

$$X^K = V_{11}^+ [B_{11}(X^{K-1}, Y^{K-1})X^{K-1} + B_{12}(X^{K-1}, Y^{K-1})Y^{K-1}] \quad (15)$$

$$Y^K = V_{22}^+ [B_{12}(X^{K-1}, Y^{K-1})^t X^{K-1} + B_{22}(X^{K-1}, Y^{K-1})Y^{K-1}]$$

donde

- $[V_{11}^+]_{N \times N} = R^{-1}(I - (N + R)^{-1}11^t)$
- $[V_{22}^+]_{R \times R} = N^{-1}(I - (N + R)^{-1}11^t)$
- 1 es el vector columna cuyos elementos son 1.
- $B_{12}(X^{K-1}, Y^{K-1})$ es una matriz cuyos elementos son:

$$b_{ik} = \begin{cases} -\frac{\delta_{ij}}{d_{ij}(X^{K-1}, Y^{K-1})} & \text{si } d_{ij}(X^{K-1}, Y^{K-1}) \neq 0 \\ 0 & \text{si } d_{ij}(X^{K-1}, Y^{K-1}) = 0 \end{cases}$$

- $B_{11}(X^{K-1}, Y^{K-1})$ es la matriz diagonal de elementos: $b_{jj} = -\sum_i b_{ij}$
- $B_{22}(X^{K-1}, Y^{K-1})$ es la matriz diagonal de elementos: $b_{ii} = -\sum_j b_{ij}$

En contraposición a los modelos de Unfolding Interno se encuentran los del Unfolding Externo.

- **Unfolding Externo:** En estos modelos, se supone que la configuración de estímulos ya es conocida (posiblemente de un análisis MDS previo que utiliza proximidades entre objetos). A partir de datos de preferencia sobre esos objetos proporcionados por unos sujetos, el Unfolding externo sitúa los puntos ideales para cada sujeto de forma que se obtenga la representación conjunta que verifica que cuanto más próximo esté un punto de objeto a un punto de sujeto más preferencia siente ese individuo hacia ese objeto.

El programa PREFMAP (PREFerence MAPping) (Carroll, 1972; 1980) realiza el Unfolding externo, empleando o bien una transformación monótona o bien una transformación lineal de los datos, para cuatro modelos diferentes:

- I. Modelo Unfolding general
- II. Modelo Unfolding ponderado
- III. Modelo Unfolding simple

IV. Modelo vectorial

Estos modelos son las cuatro “fases” del programa y forman una jerarquía, ya que cada fase es un caso especial de la fase precedente.

El **modelo vectorial** (Tucker, 1960) (**Fig. 14**) se diferencia del modelo del punto ideal en el hecho de que cada individuo no se representa a través de un punto sino que se representa mediante un vector. Con este modelo, la proyección de las posiciones de los estímulos sobre el vector de un sujeto (cuyo extremo indica la máxima preferencia) deberá reflejar las preferencias de ese sujeto. Además, el ángulo que forma el vector con cada eje o dimensión en el espacio de representación se puede interpretar como la importancia que tiene esa dimensión en las elecciones o preferencias de ese sujeto y las diferencias entre individuos según sus elecciones particulares se muestran en la diferencia de cada vector de preferencias.

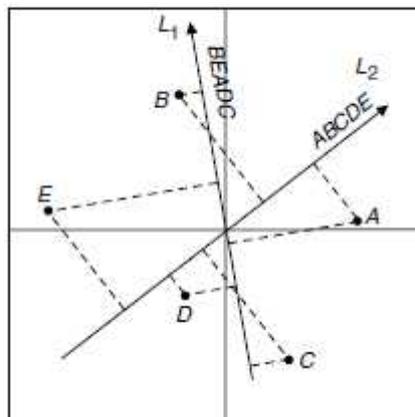


Fig. 14. Ejemplo de modelo vectorial: L1 y L2 son dos individuos (Borg y Groenen, 2005)

La expresión del modelo vectorial viene dada por:

$$\delta_{ij} = \sum_{a=1}^n w_{ia} x_{ja} + e^i \quad (16)$$

donde:

- δ_{ij} es la puntuación de la preferencia de un individuo i ($i: 1,2...R$) por un estímulo j ($j: 1,2..., N$), asignada de forma que a mayor valor de δ_{ij} menor es la preferencia.
- w_{ia} es el peso asociado al individuo i obtenido mediante regresión lineal.

- x_{ja} es la posición del estímulo j a lo largo del atributo a (a: 1,2,...,R).
- e_i es una constante aditiva única para el individuo i.

El modelo Unfolding simple, al que ya se ha hecho referencia anteriormente, supone que las preferencias vienen dadas por:

$$i_j = \sum_{a=1}^R (y_{ia} - x_{ja})^2 + e_i \quad (17)$$

donde

- y_{ia} es la coordenada del punto ideal que representa al individuo i en la dimensión a.
- x_{ja} es la coordenada del punto que representa al estímulo j en la dimensión a.

Este modelo supone que los sujetos basan sus juicios de preferencia en los mismos atributos pero difieren en cuanto a la posición de los puntos ideales. El modelo simple es el modelo circular del punto ideal ya que la preferencia de un sujeto por un objeto disminuye a medida que aumenta el cuadrado de la distancia de ese objeto al punto ideal del sujeto.

El **modelo Unfolding ponderado** (Carroll, 1980) asume que los individuos comparten los mismos ejes de referencia (atributos), pero lo que diferencia a dos individuos es el valor (o peso) que cada uno otorga a una determinada dimensión en función de lo importante que sea para él. Por lo tanto, las distancias dentro del espacio que comparten se verán afectadas en esa dimensión. Así la expresión de las preferencias viene dada por:

$$i_j = \sum_{a=1}^R W_{ia}^2 (y_{ia} - x_{ja})^2 + e_i \quad (18)$$

La solución geométrica para este modelo es elíptica, ya que cada punto ideal se encuentra en el centro de elipses concéntricas y los objetos se sitúan sobre las elipses más o menos cercanas en función de la mayor o menor preferencia. El eje más largo representa a la dimensión menos importante, pues para provocar una variación en la intensidad del atributo al que representa se debería generar un cambio grande en esta dimensión; mientras que el eje que representa a la dimensión más importante es el más corto, ya que cualquier pequeña fluctuación en esta dimensión provoca un gran cambio en el atributo correspondiente.

Por último, el **modelo general de Unfolding** viene dado por:

$$ij = \sum_{a=1}^n w_{ia}^2 (y'_{ja} - x'_{ia})^2 + e_i \quad (19)$$

donde $X^*=XT_i$, $Y^*=YT_i$ y T_i es la matriz de rotación ortogonal propia del individuo i -ésimo.

Este modelo permite a cada individuo participante la libertad implícita de elección de su propio sistema de referencia, de modo que cada uno de ellos puede rotar los ejes de referencia del espacio perceptual y otorgar pesos diferentes a las dimensiones definidas por cada eje.

Revisión del Software en Unfolding

Hasta este punto, se ha realizado una revisión conceptual del Unfolding: la clasificación según el tipo de datos y las transformaciones de los mismos y los problemas que conlleva el intento de búsqueda de una configuración óptima para individuos y estímulos.

Para continuar este trabajo, es lógico proseguir con una revisión del software disponible.

Al ser el Unfolding un caso especial de MDS, muchos de los programas que resuelven datos de MDS también resuelven Unfolding aunque, como ya se ha visto en capítulos anteriores, a través de la evolución y perfeccionamiento de cada programa se ha podido llegar a configuraciones óptimas de los datos libres de soluciones degeneradas y triviales.

A continuación se presenta una muestra del software y programas disponibles que resuelven MDS, que también están preparados para analizar datos de preferencia o Unfolding.

MINISSA

Su nombre como acrónimo proviene de *Michigan Israel Netherlands Integrated Smallest Space Analysis*, su autor es Roskam el cual en 1968 tras haber trabajado previamente con Kruskal y con Lingoés, pudo realizar una profunda revisión y comparativa de los algoritmos desarrollados por Kruskal y Carmone (1969) (M-D-SCAL) y Guttman y Lingoés (1966) (SSA, *Smallest Space Analysis*). El resultado de

esta colaboración vio la luz en un suplemento de la revista *Psychometrika* (Lingoes y Roskam, 1973) y la aparición del programa MINISSA (Roskam y Lingoes, 1970).

MINISSA y MINIRSA (*Michigan Israel Netherlands Integrated Rectangular Space Analysis*) incluyen la combinación de los programas desarrollados por Kruskal y contienen una aproximación computacional del problema de minimización basada en técnicas originales de Kruskal y Guttman que asegura la convergencia, además de la elección de la configuración inicial para evitar soluciones degeneradas aunque no las evita. Como se verá a continuación, ambos algoritmos forman parte del paquete NEWMDSX.

KYST

KYST es un programa diseñado por Kruskal y Carroll en 1969 para resolver MDS y Unfolding. Combina M-D-SCAL (el primer programa desarrollado por Kruskal para encontrar soluciones MDS) y TORSCA (Young y Torgerson, 1967). El programa y su manual de uso pueden ser descargados de la Web <http://www.netlib.no/netlib/mds/>. En sus inicios, partía de una solución inicial obtenida a través de un escalamiento clásico de Torgerson y no evitaba las soluciones degeneradas. En 1982 Borg y Bergermaier consiguieron parcialmente eliminarlas a través de un Unfolding de intervalo condicional por filas con un término independiente $a=0,5$ en la función de transformación. En 2005 Borg y Groenen en un intento por evitar las soluciones degeneradas, proponen cambiar el tipo de Unfolding anterior por uno de razón, sustituyendo el término independiente 0,5 por 0, es decir, eliminándolo.

ALSCAL

ALSCAL, acrónimo de *Alternate Least Squares Scaling* (Takane, Young y deLeeuw, 1977), es hoy en día uno de los módulos existentes en SPSS para generar soluciones MDS. Fue el primer algoritmo capaz de realizar tanto análisis métrico como no métrico según el modelo euclídeo ponderado o el simple. Perfeccionó el ajuste del modelo a los datos mediante una variante del método de los mínimos cuadrados, que es convergente, muy rápida y relativamente libre del problema de los mínimos locales. Difiere de otros programas que resuelven MDS en que minimiza el S-Stress en vez del Stress, ajustando distancias cuadradas a disparidades cuadradas. En consecuencia, en una solución ALSICAL las disimilaridades grandes se ven mucho mejor

representadas que las pequeñas. Como dato positivo se ha de señalar que además resuelve modelos para datos asimétricos, Unfolding y Datos de tres vías o entradas (a través del modelo euclídeo general o ponderado) y realiza Unfolding externo, ya que permite fijar coordenadas.

NEWMDSX

NEWMDSX es un paquete estadístico que engloba ciertos programas individuales aptos para realizar distintos tipos de MDS y Unfolding desarrollados entre 1960 y 1980. Actualmente es uno de los paquetes de más calidad para resolver problemas de MDS y es compatible con cualquier versión de Windows incluida Windows 8 (Mayo 2013), tiene una buena resolución gráfica y permite la entrada de datos en diversos formatos. A continuación se detallan brevemente los programas que contiene relativos a Unfolding:

- CANDECOMP (*CANonical DECOMPosition*) es un algoritmo propuesto por Carroll y Chang (1970) orientado a análisis de datos de N (máximo 7) vías. Proporciona soluciones a Unfolding Interno mediante la descomposición de matrices rectangulares con datos de disimilaridad o medidas de correlación. Utiliza una transformación lineal de los datos y escala de intervalo.
- INDSCAL, acrónimo de *INDividual Differences SCALing for Fitting the Euclidean Distances*, también fue propuesto por Carroll y Chang (1970), proporciona Unfolding Interno, para matrices de dos vías. Utiliza una transformación lineal de los datos y se basa en el modelo de distancia euclídea ponderada.
- MDPREF, acrónimo de *MultiDimensional PREFERence* (Chang y Carroll, 1969), proporciona Unfolding Interno para matrices de dos vías. Utiliza una transformación lineal de los datos y se basa en el modelo de vectorial.
- MINISSA y MINIRSA, ya vistos anteriormente.
- PREFMAP, acrónimo de *PREFERence MAPPING* (Carroll, 1972), es un conjunto de programas y algoritmos que proporcionan análisis tanto para Unfolding externo como interno, datos métricos y no métricos y el modelo del punto ideal o el modelo vectorial.

SMACOF-III

Tras una serie de implementaciones y desarrollos del algoritmo original SMACOF (acrónimo de *Scaling by Majorizing a COmplex Function*) de deLeeuw y Heiser (1977), la variante métrica del Unfolding se llama SMACOF-III (Heiser y deLeeuw, 1979) y la no métrica SMACOF-IIIb (Heiser, 1987), aunque desafortunadamente el código no está disponible en la actualidad. Lo interesante en este algoritmo es la configuración inicial: los estímulos (columnas de la matriz) son los centroides (baricentro o centro de masas) de aquellos individuos (filas de la matriz) que los han puntuado con la mayor preferencia. Según sus autores, esta condición se utiliza para tener una buena configuración inicial y mejores interpretaciones de la representación conjunta y, de hecho, esta idea también la llevan a cabo de manera más desarrollada DeSarbo y Rao (1984) en su algoritmo GENFOLD-2 como manera de evitar soluciones degeneradas.

GENFOLD-2

GENFOLD-2, acrónimo de *GENeral unFOLDing analysis of preference/dominance Data versión 2*, es un algoritmo desarrollado por DeSarbo y Rao (1984) no disponible aún en ningún paquete estadístico. Kim, Rangaswamy y DeSarbo (1999) hacen mención a una versión anterior GENFOLD. DeSarbo y Rao (1984) describen en su artículo una compilación de modelos Unfolding de dos entradas para datos de preferencia. Esta compilación contiene una serie de opciones como Análisis Externo o Interno, Condicional o Incondicional, la utilización de transformaciones métricas o no métricas de los datos así como los modelos general, simple y ponderado.

Utiliza una función objetivo común para todos los modelos, ponderada por los distintos valores de no-preferencia con el fin de evitar soluciones degeneradas. Para la minimización de la función objetivo, GENFOLD-2, utiliza el método alternado de los mínimos cuadrados ponderados.

Los autores de GENFOLD-2, consideran que es necesario utilizar diferentes tipos de ponderaciones en la función objetivo dependiendo de los datos de entrada (razón, intervalo u ordinal), la fiabilidad en los juicios de preferencia o la condicionalidad para evitar las soluciones degeneradas, tal y como se comentó en el apartado **(Asignar pesos)** del capítulo anterior.

La variante de este método para tres entradas o vías propuesta por DeSarbo y Carroll (1985), se revisa en el capítulo siguiente.

NEWFOLD

NEWFOLD es un algoritmo propuesto por Kim, Rangaswamy y DeSarbo en 1999 y todas las especificaciones acerca de la optimización se describen en el artículo de Kim (1990). NEWFOLD es un programa ejecutable en entorno MS-DOS. Puede manejar datos métricos y no métricos y realiza Unfolding interno y externo. La configuración inicial puede ser coherente con los estímulos, aleatoria o propuesta por el usuario; el algoritmo utiliza el método del gradiente conjugado¹⁰ para el cálculo de la configuración óptima de individuos y estímulos aunque en la práctica no proporciona buenas medidas de ajuste en los datos.

PREFSCAL

PREFSCAL, acrónimo de *PRE*ference *SCAL*ing, es un algoritmo propuesto por Busing, Heiser, Neufeglise, y Meulman en 2005 y se encuentra disponible en SPSS. La búsqueda de algoritmos exentos de soluciones degeneradas siempre ha acompañado al Unfolding, realmente es una evolución de SMACOF-IIIb de Heiser. Tal y como se comentó anteriormente, utiliza un coeficiente de penalización para prevenir soluciones degeneradas y la función de minimización es idéntica a la que se utiliza en SMACOF-IIIb, aunque utiliza un algoritmo de actualización de la configuración diferente al anterior (Busing, 2010). PREFSCAL, por lo tanto, calcula configuraciones óptimas y evita soluciones degeneradas aunque quizás el tiempo computacional puede ser más largo de lo habitual.

Un paso más: Unfolding de Tres Vías (Three-Way Unfolding)

Paralelamente al desarrollo de las técnicas para hallar configuraciones óptimas y libres de soluciones triviales y degeneradas en Unfolding clásico, algunos autores como DeSarbo y Carroll se han interesado por el desarrollo de métodos y programas aptos para el cálculo de Unfolding de datos de preferencia de tres vías.

Desde el punto de vista estadístico, una tabla de tres entradas o vías (individuos, variables, ocasiones [npxT]) puede considerarse como una sucesión de T tablas de

¹⁰El **Método del Gradiente Conjugado** es un método iterativo de optimización lineal sin restricciones. Se utiliza por ejemplo para minimizar funciones cuadráticas bajo ciertos supuestos. La idea del algoritmo es utilizar direcciones conjugadas para el descenso en la búsqueda del punto óptimo x^* . Dos vectores son conjugados si son ortogonales, es decir, si su producto escalar es cero.

dos entradas (individuos, variables [n x p]). Sin embargo, los métodos estadísticos que analizan éstas no están indicados para el tratamiento de las de tres entradas, ya que no permiten la comparación de los resultados individuales de las T tablas.

Los métodos para el análisis conjunto de varias matrices de datos fueron desarrollados inicialmente por dos escuelas, la francesa y la anglosajona, pero en la actualidad existen otras contribuciones importantes, que no pertenecen a ninguna de las dos, aunque algunas estén vinculadas más a una que a otra por la forma en que realizan el análisis.

En general, se puede agrupar los métodos para el tratamiento de datos de tres modos de acuerdo a dos objetivos:

1. Los que se centran en ajustar modelos que reproduzcan lo más fielmente posible los datos originales; es decir, están basados en la existencia de "modelos subyacentes" en la tabla de datos. Entre ellos se encuentran el modelo de Tucker (1966); el PARAFAC o *PARAllel FACtor analysis* (Harshman, 1970; Harshman y Lundy, 1984), el CANDECOMP o *CANonical DECOMPosition* (Carroll y Chang, 1970), el INDSCAL o *INDividual Differences SCALing* (Carroll y Chang, 1970) y el IDIOSCAL o *Individual Differences In Orientation SCALing* (Carroll y Chang, 1972) para diferencias individuales. Los métodos Tuckals, Tuckals2 y Tuckals3 (Kroonenberg y deLeeuw, 1980), basados en los modelos Tucker2 y Tucker3 (Tucker, 1966, 1972); y el modelo de Escalamiento de Tres Modos (*Three Mode Model*) (Kroonenberg, 1983), que es un caso particular del método Tuckals3, donde dos de los tres modos coinciden.

Estos métodos están pensados para trabajar con las matrices de datos originales, es decir, que realizan un ajuste directo de los datos y tratan a los modos en forma simétrica.

2. Los que están basados en la búsqueda de una configuración consenso "óptima", que aproxime lo máximo posible las distintas configuraciones. Entre ellos se encuentran los métodos Procrustes (Gower, 1975); el Doble Análisis en Componentes Principales introducido por Bouroche en 1975; el Statis y el StatisDual desarrollado por L'Hermier des Plantes (1976); las Meta-Componentes Principales (Krazanowski, 1979, 1982); el Análisis de Componentes Principales Comunes (Flury, 1988, 1995); el Análisis Factorial Múltiple (Escofier y Pagès, 1984, 1990); el Análisis de Co-inercia Múltiple

propuesto por Chessel y Hanafi (1996) y los Meta-Biplots desarrollados por Martín-Rodríguez (1996) y Martín-Rodríguez et al. (2001), entre otros.

Estos métodos, tratan a los modos en forma *asimétrica* y trabajan con *datos derivados* ya que el consenso lo forman a partir de las configuraciones obtenidas de algún análisis factorial o a partir de matrices de productos cruzados (matrices de distancias, similitudes o covarianzas).

Los métodos Procrustes no trabajan con *datos derivados* ya que para la construcción del consenso realizan transformaciones (rotaciones, traslaciones, etc.) de las matrices originales.

La situación que hasta este momento siempre se ha expuesto es la representación en un espacio conjunto de individuos y estímulos en función de las elecciones de los individuos hacia la colección de estímulos, pero también se puede uno preguntar qué puede pasar si estas elecciones se miden consecutivamente en diferentes situaciones (tiempo, condiciones, contextos, etc.).

Según exponen DeSarbo et al. (2009), las preferencias varían según el contexto en el que se realice la elección, simplemente porque los individuos eligen dependiendo de la situación.

Lewin (1951) ya estudió lo influenciado que era la preferencia según el contexto y argumentó que el comportamiento depende tanto de la situación como del conjunto de actitudes que rodean al individuo que toma la decisión. Esta afirmación, además, se respalda con publicaciones en torno a Psicología de Consumo, Psicología del Comportamiento y Teoría de la Decisión (Belk, 1974, 1975; Puto, 1987; Payne, Bettman y Johnson, 1992) y estudios que demuestran que la preferencia y la elección son dinámicas bajo ciertas influencias de contexto (Payne, 1982; Tversky y Simonson, 1993).

También las situaciones en estudios de preferencia pueden no ser únicamente de carácter temporal (tiempo, estaciones, meses etc.), sino que pueden ser de otro tipo como por ejemplo de carácter físico (ubicación, clima...) o social (interacción entre individuos, estados de ánimo...).

En el caso de añadir un nuevo elemento situacional, el modelo propuesto por DeSarbo y Carroll (1985) representa los individuos como puntos ideales, los estímulos como puntos en un espacio conjunto D-dimensional y los pesos asignados a cada

situación t (t: 1 ... T) tienen que ser estimados y representados en los ortantes¹¹ positivos de los espacios individuales. Estos pesos, además, explicarían la importancia otorgada a la dimensión implícita de cada situación.

El modelo de Unfolding de tres vías propuesto asume que los estímulos se representan en un espacio común, pero que están ponderados de forma diferente según la situación; así, estos pesos permiten estimar los respectivos espacios perceptuales relevantes en cada situación particular. Además, el modelo de punto ideal en esta nueva situación puede adquirir dos vertientes distintas:

- a) El modelo de punto ideal flotante, en el que se asume que la representación de los puntos ideales (individuos) varía en las diferentes situaciones, por lo tanto tienen que ser estimados de nuevo para cada situación t diferente.
- b) El modelo de punto ideal fijo en el que los puntos ideales permanecen inmóviles en todas las situaciones.

La elección del tipo de modelo de punto ideal será parte de la elección del algoritmo óptimo para encontrar soluciones en este modelo.

Modelo Ponderado de Tres Vías (*The Weighted Least Squares Algorithm for Three-Way Unfolding*)

La asunción sobre la que subyace este modelo es una generalización del propuesto por Schönemann (1970). En él se presupone que los valores de no-preferencia (inversamente relacionados con los de preferencia) se miden a través de una escala de intervalo o de razón y están relacionados linealmente con las distancias euclídeas al cuadrado, indicando así que la preferencia disminuye rápidamente conforme se desvía en cualquier dirección del punto ideal que representa a un individuo:

$$i_{jt} \cong d^2_{ijt} + \alpha t = \sum_{d=1}^v \omega_{dt} [(y^i_{td} - x^j_{td})]^2 + \alpha t \quad (20)$$

donde:

- i: 1..., R es el subíndice asociado a los individuos; j: 1..., N, el asociado a los estímulos; t: 1..., T, el asociado a las situaciones y d: 1..., D, a las dimensiones.

¹¹ En Geometría un **ortante** o **hiperortante** es el análogo en un espacio euclídeo N-dimensional a un cuadrante en el plano o un octante en 3 dimensiones.

- t_{ijt} es el valor de no-preferencia de un individuo i respecto al estímulo j en una situación t .
- d_{ijt}^2 es la distancia euclídea al cuadrado entre el individuo i y el estímulo j en la situación t .
- ω_{dt} indica el peso o importancia de la dimensión d en la situación t .
- y_{id} es la coordenada en la dimensión d del punto ideal referido al individuo i y x_{jd} es la coordenada en la dimensión d del punto referido al estímulo j .
- α_t es una constante aditiva para cada situación t .

Conocida la matriz $\delta_{ijt} = (\delta_{ijt})$ y el número D de dimensiones, DeSarbo (1978) propuso inicialmente estimar los parámetros ω_{dt} , y_{id} , x_{jd} y α_t de forma que se minimice la función de pérdida:

$$\text{Min } Z_1 = \sum_{i=1}^R \sum_{j=1}^N \sum_{t=1}^T (t_{ijt} - \delta_{ijt})^2 \quad (21)$$

con

$$\delta_{ijt} = \sum_{d=1}^D \omega_{dt} [(y_{id} - x_{jd})]^2 + \alpha_t \quad (22)$$

El algoritmo propuesto comprende un procedimiento que combina el algoritmo CANDECOMP de Carroll y Chang con un procedimiento de mínimos cuadrados alternados que estima de manera iterativa ω_{dt} , y_{id} , x_{jd} y α_t . Sin embargo, este algoritmo, además de ser bastante costoso desde el punto de vista computacional, en ocasiones también se ve afectado por soluciones degeneradas o triviales.

Por ello, DeSarbo y Carroll propusieron en 1985, al igual que ya hizo DeSarbo en 1984 con GENFOLD-2, un algoritmo modificado para generar soluciones en Unfolding de tres vías libre de soluciones degeneradas o triviales.

En primer lugar modificaron la función de pérdida (21), incorporando pesos, γ_{ijt} :

$$\text{Min } Z_2 = \sum_{i=1}^R \sum_{j=1}^N \sum_{t=1}^T \gamma_{ijt} (\delta_{ijt} - \delta_{ijt})^2 \quad (23)$$

Estos pesos w_{ijt} , pueden ser definidos de diversas maneras en función de las no-preferencias δ_{ijt} . DeSarbo y Carroll exponen que los valores que toman los pesos, dependen en cierta medida de la fiabilidad que se otorgue a los juicios de proximidad, aunque distintos autores discuten acerca de si es más adecuado tener en cuenta las distancias más grandes o las más pequeñas entre puntos. No obstante, DeSarbo resalta que el algoritmo propuesto (***The Weighted Least Squares Algorithm for Three-Way Unfolding***) permite la especificación de diferentes tipos de ponderación w_{ijt} en relación a los datos utilizados (escala de medida, procesamiento, condicionalidad, etc.) y que, en general, la función de pesos puede ser expresada como

$$w_{ijt} = \left(\frac{1}{\delta_{ijt}} \right)^p \quad (24)$$

o bien

$$w_{ijt} = \left(\frac{1}{r(\delta_{ijt})} \right)^p \quad (25)$$

donde el exponente p puede tomar cualquier valor entero, y $r(\delta_{ijt})$ representa los rangos de los δ_{ijt} para cada fila de la matriz Δ .

Algoritmo de Mínimos Cuadrados Ponderados para Unfolding de Tres Vías

El algoritmo propuesto por DeSarbo y Carroll (1985) para resolver datos Unfolding de tres vías está basado en el método de mínimos cuadrados ponderados. Consiste en estimar α_{id} , y_{id} , x_{jd} y α_t dados δ_{ijt} , y_{ijt} y D , de manera que en cada paso del algoritmo, se calcula un parámetro reteniendo los otros con un valor fijo. Permite al usuario la opción de realizar un Unfolding externo (dado $X=(x_{jd})$) o interno (estimando X), al igual que permite la elección de datos de intervalo o de razón y la elección del modelo de punto ideal fijo y_{id} o flotante y_{itd} para representar los individuos.

El algoritmo, que se expone a continuación, considera las siguientes etapas:

1. Se obtienen las estimaciones iniciales de W , X , Y y r . Estas estimaciones pueden obtenerse aleatoriamente, pueden venir dadas por el usuario, o pueden obtenerse mediante CANDECOMP (Apéndice A, DeSarbo y Carroll, 1985). Otra posibilidad es obtener estas estimaciones promediando la matriz de tres

vías Δ para generar una matriz de dos vías que se utiliza para obtener X e Y mediante un Unfolding de dos vías, con W matriz de unos.

Se fija un contador IT (inicialmente igual a 0)=IT+1.

2. Si se opta por un análisis interno, se estima X utilizando el método del gradiente conjugado y fijando el resto de parámetros. Para ello, se calcula la derivada parcial de la función de pérdida ponderada (23) respecto de x_{jd} :

$$\frac{\partial J^2}{\partial x_{jd}} = 4 \sum_t \sum_l w_{ltd} [(\delta_{ljt} - \bar{\delta}_{ljt})(y_{ld} - x_{ld})] \quad (26)$$

De este modo se puede construir el vector gradiente ∇X , con las sucesivas derivadas parciales, colocadas en un vector de dimensión $1 \times N \times D$:

- a) Se establece el contador MIT=1 para iterar en la búsqueda del óptimo en X.
- b) Se fija la siguiente dirección $S^{(1)} = -\nabla X^{(1)}$ en la 1ª iteración
- c) Se encuentra $X^{(2)}$ de manera que

$$X^{(2)} = X^{(1)} + u^{(1)} S^{(1)} \quad (27)$$

donde $u^{(1)}$ es el valor del paso óptimo en la dirección $S^{(1)}$, obtenido por interpolación cuadrática. Se fija el contador MIT=2

- d) Se calcula $X^{(MIT)}$ y se fija el valor

$$S^{(MIT)} = -\nabla X^{(MIT)} + \frac{\nabla X^{(MIT)T} \nabla X^{(MIT)}}{\nabla X^{(MIT-1)T} \nabla X^{(MIT-1)}} S^{(MIT-1)} \quad (28)$$

- e) Se calcula el paso óptimo $u^{(MIT)}$ en la dirección $S^{(MIT)}$ y se actualiza:

$$X^{(MIT+1)} = X^{(MIT)} + u^{(MIT)} S^{(MIT)} \quad (29)$$

- f) Si $X^{(MIT)}$ es el óptimo buscado entonces el bucle para; si no, se iguala MIT a MIT+1 y se vuelve a d). Para parar en f), se requiere una de estas tres condiciones:

- I. MIT \geq máximo de iteraciones definidas por el usuario
- II. $\|X^{(MIT)}\| <$ criterio de tolerancia (TOL) definido por el usuario

$$\text{III. } |Z_2^{(MIT-1)} - Z_2^{(MIT)}| < \text{TOL}$$

3. Se estima la matriz Y de una manera similar a la utilizada en 2.
4. Se estiman la matriz de ponderaciones W y los parámetros aditivos α . Para ello, en esta etapa se mantienen fijos todos los parámetros excepto ω_{dt} , que se estima a partir de la matriz $\Delta = (\delta_{ijt})$ de dimensión $T \times (RN)$ y a partir de la matriz $P = (p_{ijd})$ con $p_{ijd} = \frac{(y_{id} - x_{jd})^2}{\alpha}$ de dimensión $(RN) \times D$, como :

$$W^{est} = \Delta P (P^t P)^{-1} \quad (30)$$

Además, para estimar las constantes α_t es suficiente con añadir una columna de unos en la matriz P anterior.

5. Se para, si se cumplen cualquiera de estas dos condiciones:
 - I. $|Z_2^{(IT-1)} - Z_2^{(IT)}| < \text{TOL}$
 - II. IT = número máximo de iteraciones definido por el usuario.
6. Si el usuario elige la opción del punto ideal fijo (la representación de los individuos no cambia con las situaciones), éstos tienen que ser ajustados por ciertos coeficientes ponderados según la situación ω_{dt} . Primero, se normaliza (de forma que su suma de cuadrados sea igual a la unidad) la matriz de dimensión $(R+N) \times D$ obtenida poniendo X e Y una detrás de la otra y se redefine la matriz de ponderación situacional W de manera que se refleje la varianza entre cada situación.
7. Si se ha elegido la opción del punto ideal flotante, entonces se asume que las coordenadas de los puntos ideales cambian en función de las situaciones. Se considera que

$$i_{jt} = a_{it} d_{ijt}^2 + b_{it} + e_{ijt} \quad (31)$$

donde

- d_{ijt}^2 es la distancia euclídea al cuadrado entre el individuo i y el estímulo j en la situación t = $\sum_{d=1}^D \omega_{dt} [(y_{id(t)} - x_{jd})^2]$

- $y_{id(t)}$ es la coordenada en la dimensión d del punto ideal referido al individuo i en cada situación t.
- x_{jd} es la coordenada en la dimensión d del punto referido al estímulo j.
- a_{it} es una constante multiplicativa referida al individuo i en cada situación t.
- b_{it} es una constante aditiva referida al individuo i en cada situación t.
- e_{ijt} es el término de error.

Si en (31) se elimina el término de error, se sustituye la igualdad por \cong y se introducen pesos γ_{ijt} para evitar soluciones degeneradas se obtiene:

$$\delta_{ijt} \cong \gamma_{ijt} a_{it} \sum_{d=1}^D \omega_{dt} [(y_{id(t)} - x_{jd})^2 + b_{it}] \cong \gamma_{ijt} a_{it} \sum_{d=1}^D \omega_{dt} y_{id(t)}^2 - 2 \gamma_{ijt} a_{it} \sum_{d=1}^D \omega_{dt} y_{id(t)} x_{jd} + \gamma_{ijt} a_{it} \sum_{d=1}^D \omega_{dt} x_{jd}^2 + \gamma_{ijt} b_{it} \quad (32)$$

Si se toma $z_{it} = a_{it} \sum_{d=1}^D \omega_{dt} y_{id(t)}^2 + b_{it}$ y $q_{idt} = -2a_{it} y_{id(t)}$ la expresión anterior queda

$$\delta_{ijt} \cong \gamma_{ijt} a_{it} \sum_{d=1}^D \omega_{dt} x_{jd}^2 + \gamma_{ijt} \sum_{d=1}^D q_{idt} \omega_{dt} x_{jd} + \gamma_{ijt} z_{it} \quad (33)$$

y se puede estimar los puntos ideales como $y_{id(t)} = -\gamma_{ijt} / 2 \left(\frac{q_{idt}}{a_{it}} \right)$

El algoritmo propuesto es eficiente desde el punto de vista computacional y también reduce el riesgo de la aparición de soluciones degeneradas en la configuración conjunta.

Para comprobar el rendimiento del algoritmo, sus autores efectuaron un análisis de Monte Carlo con una combinación aleatoria de factores, variando los valores D, T, R (individuos), N (estímulos), las funciones ponderadas γ_{ijt} , el error en Δ_{ijt} , la configuración inicial, el tipo de puntos ideales (fijos o flotantes), la escala de medida de los datos y el tipo de análisis (externo o interno).

Los 10 factores resultantes y sus niveles respectivos se muestran en (Tabla 1); además se usó un diseño factorial fraccionado¹² $3^8 2^3$ (Addelman, 1962) para estimar los

¹²En estadística, un **experimento factorial completo** es un experimento cuyo diseño consta de dos o más factores, cada uno de los cuales posee distintos valores o niveles, cuyas unidades experimentales cubren todas las posibles combinaciones de esos niveles en todo los factores. Este tipo de experimentos permiten el estudio del efecto de cada factor sobre la variable respuesta, así como el efecto de las interacciones entre factores sobre dicha variable. Si el número de combinaciones en un diseño factorial completo es demasiado alto para su procesamiento, puede optarse por un **diseño factorial fraccional**, en el que se omitan algunas de las combinaciones posibles.

efectos principales, con un total de 27 pruebas que resultaron de la combinación aleatoria de los 10 factores y sus niveles respectivos, de modo que cada una de las pruebas contuviera una combinación de factores única que no se repitiese en otra distinta.

Factores		Niveles (Códigos)		
1	Dimensiones D	1(0)	2(1)	3(2)
2	Situaciones T	2(0)	3(1)	4(2)
3	Individuos I	10(0)	15(1)	20(2)
4	Estímulos J	10(0)	15(1)	20(2)
5	Pesos	Ninguno	$\left(\frac{1}{r(\delta_{ijt})}\right)$	$\left(\frac{2}{r(\delta_{ijt})}\right)^2$
6	Error en los datos	Ninguno	15 %	30 %
7	Conf. inicial	Aleatoria(0)	Acercamiento(1)	CANDECOMP(2)
8	Puntos Ideales	Fijos(0)	Flotantes(1)	
9	Escala	Intervalo(0)	Razón(1)	
10	Tipo Unfolding	Interno(0)	Externo(1)	

Tabla 1. Factores Independientes y Niveles definidos para el Análisis Monte Carlo en Unfolding de Tres Vías (Fuente DeSarbo y Carroll, 1985)

De esta manera, con cada uno de los niveles mostrados, se generaron de manera aleatoria a través de una distribución uniforme las matrices W , X , Y y L y conforme a la expresión (20) se calculó la matriz de datos \hat{W} , teniendo también en cuenta otras consideraciones como el error, la función de pesos, los puntos ideales o el tipo de Unfolding.

Se registraron las siguientes medidas para valorar el ajuste de los datos, siendo la medida sobre W la que mejor ajuste proporcionó:

1. El número de iteraciones necesario para la convergencia.
2. La suma de cuadrados promedio (sobre las dimensiones) tenida en cuenta entre W y \hat{W} .
3. La varianza media (sobre las dimensiones) tenida en cuenta entre X y \hat{X} .
4. La varianza media (sobre las dimensiones) tenida en cuenta entre Y e \hat{Y} .
5. La varianza media (sobre las dimensiones) tenida en cuenta entre

$$L = \begin{bmatrix} X \\ Y \end{bmatrix} \text{ y } L = \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix}.$$

Se ajustaron también varios análisis de regresión múltiple (uno por cada medida de ajuste anterior como variable dependiente y como variables independientes en formato *dummie*¹³ los niveles de cada factor de **(Tabla 1)**, pero ninguno de ellos proporcionó resultados significativos (en términos de los test F o prueba de errores).

Respecto de las variables incluidas en cada modelo, el modelo de “acercamiento” fue significativo para las medidas 3, 4 y 5, pero con un coeficiente negativo, lo que indica que este tipo de modelos como configuración inicial no es el más adecuado. Y respecto del Unfolding Externo apareció como significativo para la medida 3, lo cual es lógico pues en este tipo de Unfolding se ha fijado previamente las coordenadas de X.

Como conclusión, los autores evidencian que el algoritmo propuesto para generar soluciones para el Unfolding de tres vías es robusto respecto de los cambios en los valores de T, D, R, N, el error en los datos y otros factores tenidos en cuenta en **(Tabla 1)**, aunque tal y como se ha comentado, el algoritmo es sensible a la configuración inicial, sugiriendo que CANDECOMP sería la más adecuada para reducir el riesgo de alcanzar soluciones óptimas poco adecuadas y por ende soluciones degeneradas. Sin embargo, también resaltan el hecho de que existen limitaciones como, por ejemplo, que la utilización del diseño factorial fraccionado puede no detectar la interacción entre los efectos de los factores y que la investigación debe centrarse en diseños de mayor tamaño y un incremento en el número de niveles de cada factor y también en la función de pesos.

Desarrollos Recientes: Unfolding de Tres Vías y Análisis Cluster

Adentrándose ya en el Siglo XXI, de nuevo DeSarbo, Atalay y Blanchard (2009), inician una nueva andadura y proponen nuevos retos en el Unfolding de Tres Vías.

Dentro del Análisis Multivariante es conocido el Análisis Cluster como una manera rápida y efectiva de representar la homogeneidad de una muestra respecto de una serie de características. Numerosos autores han propuesto a lo largo del tiempo la combinación de ciertas técnicas de Análisis Multivariante junto al Análisis Cluster. La combinación de MDS junto al Análisis Cluster aparece ya en la literatura en los 90 desde dos puntos de vista diferentes.

Según el primer punto de vista, DeSarbo et al. (1994) proponen la combinación de los Modelos de Clases Latentes junto al MDS (en adelante, LCMDS) recogiendo lo ya propuesto anteriormente por otros autores como Tucker (1960), Slater (1960) o

¹³Una variable **dummie**, es una variable artificial, de formato binario (0,1) que se crea a partir de una variable nominal. En regresión se crean tantas variables *dummie* como categorías tiene la variable original menos 1.

Coombs (1964). En este caso, en el análisis LCMDS se estiman clusters a través del Método de Máxima Verosimilitud (algoritmos EM), en vez de estimar individuos a través de vectores o puntos ideales, reduciendo por tanto el número de parámetros en el cálculo. La revisión histórica de esta aproximación a través del LCMDS muestra por ejemplo el desarrollo de DeSarbo et al. (1990) de un modelo vectorial LCMDS para datos normalmente distribuidos; DeSarbo et al. (1991) extienden lo anterior a un modelo de punto ideal ponderado; DeSoete y Winsberg (1993) y DeSoete y Heiser (1993) extienden los dos modelos anteriores ajustando restricciones lineales a las coordenadas en los estímulos; Böckenholt y Böckenholt (1991) desarrollan sendos modelos (simple y ponderado) de punto ideal para datos binarios; DeSarbo et al. (1995) desarrollan un modelo LCMDS vectorial para datos distribuidos según la función de Dirichlet¹⁴; Chintagunta (1994) desarrolla un modelo LCMDS vectorial para datos multinomiales y, para finalizar, Wedel y DeSarbo (1996) extienden su modelo LCMDS vectorial para datos distribuidos mediante la familia de funciones exponenciales.

Desde el otro punto de vista, la combinación de ambos métodos (representación y clasificación de individuos y sus elecciones de preferencia) también ha llevado a numerosos autores como Heiser, Groenen, etc. a proponer aproximaciones determinísticas¹⁵ a través del MDS (Heiser y Groenen, 1997; Kiers et al., 2005), el Escalamiento Óptimo (Van Buuren y Heiser, 1989) y el Análisis de Componentes Principales (DeSoete y Carroll, 1994; Vichi y Kiers, 2001). Para datos de tres entradas, existen ciertas aproximaciones de fusión entre el MDS y el análisis cluster (Meulman y Verboon, 1993; Carroll y Chaturvedi, 1995); sin embargo, DeSarbo no encontró referencia previa en la literatura a la existencia de algún procedimiento que calculase conjuntamente soluciones MDU o Unfolding de Tres Vías junto con análisis cluster. De este modo, DeSarbo, Atalay y Blanchard (2009) proponen un procedimiento mediante el cual es posible la representación conjunta de datos de preferencia de dos o tres vías o entradas junto a un análisis cluster mediante el modelo del punto ideal ya conocido. Los autores resaltan que no es necesario asumir hipótesis paramétricas

¹⁴La **función de Dirichlet** es una función real de variable real que no es continua en ningún punto de la recta. Se define como $f(x) = \begin{cases} 1 & \text{si } x \text{ es racional} \\ 0 & \text{si } x \text{ es irracional} \end{cases}$

¹⁵Los procesos llamados **determinísticos** son aquellos en los que conociendo las condiciones iniciales siempre siguen el mismo curso y producen el mismo resultado final, es decir, no existen elementos aleatorios. El estado final siempre es el mismo dado unas mismas condiciones iniciales.

como requieren los modelos LCMDS, los clusters no se solapan y el procedimiento es válido tanto para Unfolding clásico como para el de tres vías.

Solamente hay que hacer un apunte, hasta ahora el procedimiento tradicional ofrecía la representación conjunta de individuos como puntos ideales y los estímulos de manera que previamente había que realizar un cálculo de coordenadas (de cada individuo y estímulo) en el nuevo sistema de referencia más otra serie de parámetros específicos del modelo elegido.

En el modelo propuesto por DeSarbo et al. (2009) la representación conjunta se basa en la estimación simultánea de las coordenadas de los estímulos, los puntos ideales de cada cluster en cada situación y la matriz de pertenencia de los individuos a cada cluster estimado, con los mismos datos que si se realizara un Unfolding tradicional.

El modelo propuesto continúa el modelo de Unfolding de Tres Vías de DeSarbo y Carroll (1985) ya explicado en el apartado anterior. Este nuevo modelo se sustenta sobre numerosa literatura acerca de Psicología y Teoría de la Conducta y la Decisión.

Modelo Cluster en Unfolding de Tres Vías

El objetivo principal es proporcionar un modelo válido para estimar simultáneamente la representación conjunta de los estímulos y los puntos ideales asociados a cada cluster en diferentes situaciones, así como la configuración de clusters en cada una de ellas. Por lo tanto el modelo completo se podría describir del siguiente modo:

$$i_{ijt} = \sum_{s=1}^S p_{is} \sum_{d=1}^D \omega_{dt} [(x_{jd} - y_{sdt})]^2 + b_t + \varepsilon_{ijt} \quad (34)$$

donde:

- i : 1...R individuos; j : 1...N estímulos; t : 1...T situaciones; s : 1...S clusters y d : 1...D dimensiones.
- Δ_{ijt} es el valor de no-preferencia de un individuo i respecto al estímulo j en una situación t .
- ω_{dt} es el coeficiente de ajuste para la dimensión d en la situación t .
- y_{sdt} es la d -ésima coordenada del punto ideal asociado al cluster s en la situación t .
- x_{jd} es la d -ésima coordenada del punto correspondiente al estímulo j .
- b_t es una constante aditiva para cada situación t .
- ε_{ijt} es el error
- $p_{is} = \begin{cases} 1 & \text{si el individuo } i \text{ se clasifica en el cluster } s \\ 0 & \text{en caso contrario} \end{cases}$

Además como hipótesis de partida se asume que, o bien $p_{is} \in \{0,1\}$ y $\sum_{s=1}^S p_{is} = 1 \forall i$, si los clusters no se solapan, o bien $0 < \sum_{s=1}^S p_{is} \leq S$ si los clusters se solapan.

La expresión anterior (34) es prácticamente la misma que en modelo de DeSarbo y Carroll (1985) (20) salvo en la generalización para la inclusión del análisis cluster.

Algoritmo de estimación

Dados $\Delta = \Delta_{ijt}$ y los valores S y D, se pretende estimar $W = \omega_{dt}$, $b = b_t$, $P = p_{is}$, $X = x_{jd}$ e $Y = y_{sdt}$, de forma que se minimice la función:

$$= \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^T q_{ijt} [(\Delta_{ijt} - \bar{\Delta}_{ijt})]^2 \quad (35)$$

donde $\hat{\Delta}_{ijt} = \sum_{s=1}^S p_{is} \sum_{d=1}^D \omega_{dt} [(x_{jd} - y_{sdt})]^2 + b_t$ y q_{ijt} es una función de ponderación preespecificada, que se introduce para prevenir soluciones degeneradas (DeSarbo y Carroll, 1985; DeSarbo y Rao, 1984).

Los autores idean un procedimiento de estimación a través de un algoritmo de mínimos cuadrados alternados en cuatro etapas; sin embargo, dada la naturaleza no lineal del modelo, la convergencia hacia el óptimo global no estaría garantizada a priori. Las etapas de cálculo son las siguientes:

ETAPA1: Estimar W y b

Se define el parámetro

$$\hat{\Delta}_{ijt} = \sum_{s=1}^S p_{is} [(x_{jr} - y_{sdt})]^2 \quad (36)$$

siendo

- L_t la matriz de dimensión $(R^*N) \times D$ formada por los elementos concatenados de $\hat{\Delta}_{ijt}$ para cada $t=1, \dots, T$.
- $K_t = (1/L_t) \times (Q_t)^{1/2}$; siendo la matriz Q_t de dimensión $(R^*N) \times (D-1)$ formada por columnas que contienen los vectores concatenados cuyos elementos son los q_{ijt} .
- El vector unitario (traspuesto) 1^t de dimensión $(R^*N) \times 1$.
- El vector m_t de dimensión $(R^*N) \times 1$ formado por la concatenación de los elementos de $\sum_{j=1}^J q_{ijt} \times (q_{ijt})^{1/2}$, para $t=1, \dots, T$.

Conforme a la notación de DeSarbo y Carroll (1985) para el modelo de Unfolding de tres vías métrico, se considera:

$$\begin{pmatrix} \mathbf{b}_t \\ \mathbf{w}_t \end{pmatrix} = (\mathbf{K}_t^T \mathbf{K}_t)^{-1} \mathbf{K}_t^T \mathbf{m}_t \quad (37)$$

La expresión anterior se emplea para obtener los valores b_t y ω_{dt} en cada situación t , si se fijan valores de P, X e Y . Los autores además comentan la opción de emplear ciertas restricciones en W para que no haya posibilidad de que contenga valores negativos, ya que las ponderaciones negativas son difíciles de interpretar en los modelos Unfolding de tres vías.

ETAPA2: Estimar X e Y

En esta segunda etapa, se fijan los valores \mathbf{b}, \mathbf{W} y P y se estiman X e Y mediante el método del gradiente conjugado (Apéndice B, DeSarbo y Carroll, 1985). Las derivadas parciales de la función Φ (35) respecto a x_{jd}, y_{sdt} son:

$$-\frac{\partial \Phi}{\partial x_{jd}} = -4 \sum_{i=1}^n \sum_{t=1}^T q_{ijt} [(\Delta_{ijt} - Z_{ijt}) \sum_s p_{ts} \omega_{dt} [(x_{jd} - y_{sdt})]] \quad (38)$$

$$-\frac{\partial \Phi}{\partial y_{sdt}} = +4 \sum_{i=1}^n \sum_{j=1}^J q_{ijt} [(\Delta_{ijt} - Z_{ijt}) p_{ts} \omega_{dt} [(x_{jd} - y_{sdt})]] \quad (39)$$

La idea general de este método es utilizar direcciones conjugadas (ortogonales) para el descenso en la búsqueda del punto óptimo; de esta manera, localiza mínimos globales de la función sin necesidad de un número alto de iteraciones.

ETAPA3: Estimar P

Para estimar la matriz P de pertenencias, se observa que la función (35) contiene la suma de errores o discrepancias al cuadrado sobre t (situaciones), j (marcas) e i (individuos). De esta manera se puede expresar como la suma sobre i de una "partición" de funciones objetivo i :

$$l = \sum_{t=1}^T \sum_{j=1}^N q_{ijt} \left(\Delta_{ijt} - \sum_{s=1}^S p_{is} \sum_{d=1}^D \omega_{dt} (x_{jd} - y_{sdt})^2 - bt \right)^2 \quad (40)$$

ya que sólo los parámetros p_{is} y q_{ijt} llevan indexado el contador de individuos i . Debido a la separabilidad de la función con respecto a i , se puede minimizar haciendo lo propio con cada l para cada individuo i . Gracias a esta propiedad de separabilidad el procedimiento que se describe genera un conteo de $2^S - 1$ clusters (si se solapan) o S (si los clusters son disjuntos) para obtener un óptimo global de P condicionado a los valores fijos de X , Y , W y b .

ETAPA4: Evaluación de la convergencia

El algoritmo que proponen DeSarbo et al. (2009) estima los parámetros necesarios de manera cíclica e iterativa, evaluando las expresiones (37)-(39) sobre el total de individuos N , de forma que se minimiza el error cuadrático en (35). La etapa 2, proporciona mediante el gradiente conjugado al menos óptimos locales para X e Y condicionados a que el resto de parámetros queden fijos. Las etapas 1 y 3 proporcionan óptimos globales para W , b y P condicionados a que el resto de parámetros queden fijos. Según los autores, el proceso en sí no garantiza que se pueda obtener una solución óptima y recomiendan cinco ejecuciones del algoritmo por cada solución estimada en D (dimensiones) y S (número de clusters) para después poder comparar cuál es la que se aproxima mejor a la solución ideal.

Dado que la suma de errores cuadráticos en (35) depende en gran medida del tamaño de la matriz de datos (y, por tanto, de I , J y T), de la escala de medida de las preferencias, etc., los autores proponen dos medidas para evaluar la convergencia de la solución final, aparte de la selección de los valores de S y D en el modelo final:

- I. Si la medida de los datos está en escala intervalo, se utiliza el VAF (*Variance Accounted For*) cuya expresión es:

$$VAF = 1 - \frac{\sum_t \sum_j q_{ijt} [(\Delta_{ijt} - \bar{\Delta}_{jt})]^2}{\sum_t \sum_j q_{ijt} [(\Delta_{ijt} - \bar{\Delta}_{jt})]^2} \quad (41)$$

- II. Si la medida de los datos está en escala de razón se utiliza el SSAF (*Sum of Squares Accounted For*) cuya expresión es:

$$SSAF = 1 - \frac{\sum_{i=1}^I \sum_{j=1}^J q_{ijt} \left[(\Delta_{ijt} - \widehat{\Delta}_{ijt}) \right]^2}{\sum_{i=1}^I \sum_{j=1}^J q_{ijt} \Delta_{ijt}^2} \quad (42)$$

donde $\widehat{\Delta}_{ijt} = \sum_{s=1}^S \hat{p}_{is} \sum_{d=1}^D \widehat{\omega}_{dt} [(\hat{x}_{jdt} - \hat{y}_{sdt})]^2 + \hat{b}_t$ (calculado con la estimación de los parámetros obtenidos con el modelo anterior) y $\bar{\Delta}_{..t} = \frac{\sum_{i=1}^I \sum_{j=1}^J \Delta_{ijt}}{R \times N}$.

A diferencia de lo que ocurre con las soluciones de los modelos Unfolding de dos vías, las soluciones obtenidas mediante el procedimiento propuesto para el modelo cluster de Unfolding de tres vías no pueden rotarse, aunque sí pueden reflejarse.

Según los autores, el algoritmo propuesto permite al usuario seleccionar una serie de opciones en el modelo:

1. Puede analizar datos de dos o tres entradas.
2. La matriz ponderada W se convierte en un vector en el caso de una única situación y dos entradas.
3. Permite la opción de estimar tanto puntos fijos como puntos flotantes.
4. Estima clusters disjuntos o solapados.
5. Se pueden aplicar restricciones en la matriz ponderada W, para que todos sus elementos sean positivos o iguales a cierta constante.
6. Permite realizar Unfolding Externo e Interno.
7. Además brinda la posibilidad de seleccionar una variedad de opciones iniciales (valores iniciales aleatorios o racionales¹⁶) para obtener las estimaciones iniciales de X así como para la estimación inicial de la matriz de pertenencias P (un análisis de K-Medias, media de los datos de preferencia o los valores de los puntos ideales).

Debido a la cantidad de modelos disponibles, para comprobar el rendimiento del algoritmo, efectuaron un análisis Monte Carlo con una combinación aleatoria de 9 factores con 2 niveles cada uno (**Tabla 2**).

¹⁶Una configuración inicial **racional** es aquella que se obtiene a través del modelo clásico de Torgerson de escalamiento multidimensional.

Factores		Niveles (Códigos)	
1	Individuos I	50(0)	100(1)
2	Estímulos J	10(0)	20(1)
3	Situaciones T	3(0)	6(1)
4	Clusters S	3(0)	5(1)
5	Dimensiones D	2(0)	3(1)
6	Error en los datos	10 %	20 %
7	Puntos Ideales	Fijos Y(0)	Flotantes Y(1)
8	Tipo de Cluster	Disjuntos(0)	Solapados(1)
9	Configuración Inicial	Aleatoria(0)	Racional(1)

Tabla 2. Factores Independientes y Niveles definidos para el Análisis Monte Carlo en Unfolding de 3 Vías y Cluster (Fuente DeSarbo et al., 2009)

Tal y como se ha descrito en DeSarbo y Carroll (1985), considerando un diseño factorial fraccionado 2^9 para estimar los efectos principales, con un total de 32 pruebas resultado de la combinación aleatoria de los 9 factores y sus niveles respectivos, se genera una matriz ortogonal cuyas filas contienen una combinación de factores única que no se repite en otra distinta y que contiene información de los diferentes modelos a examen, que se aplicará para evaluar el modelo asociado a cada experimento.

Al carecer de datos reales con los que validar el modelo, los autores crearon los datos artificialmente, utilizando distribuciones de probabilidad conocidas.

Para generar la matriz de pertenencias P, en el caso de la opción de los clusters solapados, los individuos fueron asignados a cada cluster $s:1...S$ a través de una distribución de Bernoulli ($p=0,5$) y en el caso de clusters disjuntos se generaron S valores aleatorios (a través de una distribución uniforme) por cada individuo, de manera que cada individuo fue asignado al cluster con un valor más alto; las coordenadas de los estímulos se generaron a través de una distribución Normal $N(0,1)$ y después se estandarizaron por columnas; las coordenadas de los individuos (puntos ideales) en cada cluster y en cada situación también se generaron a través de una distribución Normal $N(0,1)$; los pesos se generaron a través de funciones de distribución uniformes (4,7) y las constantes aditivas mediante una distribución Normal $N(0,1)$. Con todos estos datos, se construyeron los valores de no-preferencia a los que se les añadió errores obtenidos mediante distribuciones normales en las que se varió el segundo parámetro (varianza) de modo que se obtuviese la tasa deseada de varianza añadida sobre la varianza total.

Con el fin de poder englobar toda la casuística del modelo (datos de entrada, parámetros, coste computacional y clasificaciones posibles dentro de los clusters) y

para asegurar un rendimiento óptimo, se registraron las siguientes medidas de ajuste en el diseño:

1. La medida de ajuste de la configuración conjunta (X, Y), que se calculó mediante la media del VAF (*Variance Accounted For*) entre el espacio original y el obtenido en la solución.
2. La medida del coste computacional, que se calculó a través del número de iteraciones realizadas en los pasos (37)-(39) del algoritmo.
3. El tiempo CPU, que se registró en segundos.
4. El VAF del test de convergencia.
5. La recuperación de los clusters, que se analizó mediante el cálculo de las diferencias entre cada celda de las matrices de pertenencia original y final; esto es, mediante un coeficiente de concordancia que representa el número de coincidencias dividido por el número total de celdas de la matriz.
6. La medida en el caso de los pesos, que se realizó utilizando la media del SSAF calculada entre las matrices de ponderación original y final.
7. La medida del ajuste del vector b de constantes, que se calculó mediante la raíz cuadrada del error medio entre los valores finales y los de la configuración original del diseño.

Con cada una de estas medidas como variable dependiente, se realizaron varios análisis de regresión múltiple (uno por cada medida) siendo las variables independientes los niveles de cada factor de (Tabla 4) en forma *dummie*, con el fin de evaluar el modelo en base a las medidas de ajuste. La configuración conjunta (X, Y) se ajustó de manera correcta con los 9 factores del experimento y se encontró que la matriz de ponderaciones W podía verse afectada por el tamaño (número de situaciones o de dimensiones) de manera que cuantos más parámetros se quiera estimar, más afectado se ve el cálculo de la matriz W. Del mismo modo, el coste computacional (iteraciones y tiempo CPU) crece cuando se añaden más dimensiones, situaciones y número de clusters, así como si los clusters pueden solaparse. El VAF se ve afectado (a peor) cuando se incrementa el error, sin embargo alcanza mejor resultado cuando la configuración inicial es racional.

En resumen, los autores concluyen que los resultados son estimulantes y proponen nuevos caminos para la investigación y su desarrollo:

- i. Reconstruir el modelo de forma que se puedan añadir restricciones lineales en los estímulos, con el fin de mejorar la interpretación de la configuración conjunta.
- ii. Extender el modelo a datos no métricos, ya que según los autores una de las limitaciones del modelo propuesto es la aplicación de un procedimiento métrico a datos no-métricos de preferencia (ordinales). Puesto que existe suficiente literatura al respecto en Unfolding (de dos entradas) y que los resultados no difieren en exceso del tipo métrico, es posible la traslación a este modelo teniendo en cuenta además las aproximaciones de los últimos años para evitar las soluciones degeneradas que se han revisado además en este trabajo.
- iii. Debido al carácter determinístico del modelo, es necesario investigar más profundamente acerca de la elección en términos del número de dimensiones y clusters para abordar la solución final.
- iv. La última de las propuestas realizadas, implica la posibilidad de mejorar el coste computacional en el sentido de que no se tenga que penalizar el tamaño de los datos (número de individuos, estímulos, dimensiones o clusters) aun cuando ni el MDS ni el Unfolding son técnicas apropiadas para analizar gran cantidad de datos. Además, es necesario continuar en la búsqueda de soluciones a ciertos problemas asociados con los algoritmos de estimación de mínimos cuadrados alternos como son, por ejemplo, los problemas con las soluciones óptimas.

Resultados

A través de la revisión de la literatura presentada en este trabajo y también mediante una búsqueda adicional, se encuentran una serie de aplicaciones del Unfolding y Unfolding de tres vías, en las cuales los autores exponen la posibilidad de aplicación de las técnicas propuestas a ciertos conjuntos de datos.

Tal y como se ha hecho referencia a lo largo de este trabajo, el Unfolding como técnica de AM surge dentro de la Psicología, y se encuentran numerosas aplicaciones en este ámbito (Green y Rao, 1972; DeLeeuw, 2004; VanDeun, 2007).

Se encuentra un importante porcentaje de ejemplos de aplicación en Marketing (DeSarbo y Rao, 1984; Chatterjee y DeSarbo, 1992; DeSarbo et al., 1987), Política (Poole, 1984; DeLeeuw, 2004; Wang et al., 1975) o Sociología (VanDeun, 2007).

En España, aunque no es una técnica ni popular ni muy utilizada, cabe destacar las aportaciones realizadas por el equipo de la Universidad de Salamanca en el ámbito de la Ecología (Férrandez et al., 2011) y por el Dr. José Fernando Vera de la Universidad de Granada, que ha realizado interesantes aportaciones desde el punto de vista heurístico (Vera et al., 2009).

Aplicaciones del Unfolding de Tres Vías

La revisión de la literatura aporta en muchos casos numerosos ejemplos del Unfolding de tres vías en el mundo del Marketing y del Análisis de Mercado que es donde hasta ahora hay un mayor número de ejemplos de aplicación (DeSarbo y Carroll, 1985; DeSarbo et al., 2009; Park et al., 2012), Marketing y Geografía (Balabanis et al., 2004), Psicología (DeSarbo y Carroll, 1985), Psicología Clínica (Leach et al., 2001), Pedagogía (Kim et al., 1999) o Sociología y Política (De Rooij y Heiser, 2000).

DeSarbo (1985) hace hincapié que puede aplicarse de manera muy variada dentro del Marketing, como por ejemplo para investigar qué tipos de programas de televisión prefiere el público infantil en diferentes situaciones (hora del día, por quién están acompañados en ese momento...) o también para analizar las preferencias de un grupo de personas a la hora de elegir ciertas tiendas para realizar compras en diversas situaciones (regalos de Navidad, regalos de cumpleaños...). Otro apunte importante que hace el autor del modelo de tres vías es que éste no debería ceñirse sólo a un análisis meramente de situación, sino que podría aplicarse por ejemplo para investigar si existe acuerdo en la publicidad de ciertos productos evaluando mediante una muestra de individuos cuál es el anuncio más adecuado para cada uno de los

productos anunciados o también para examinar la percepción de las personas sobre ciertos personajes famosos que prestan su imagen a varios productos o marcas, suponiendo que se quiere conocer cuál de los productos se asocia mejor con un personaje u otro.

Sugiere también que dentro del Marketing y la investigación de mercados hay todavía muchos campos en los que podría ser aplicado con éxito, por ejemplo, en optimización de precios: se puede utilizar el modelo para evaluar las preferencias de los consumidores sobre diferentes marcas bajo supuestos de precios. Dentro del posicionamiento de productos o marcas, se puede aplicar para realizar análisis de preferencias de los consumidores hacia diferentes productos bajo ciertas combinaciones de atributos o características, incluso se podría examinar el porqué de los cambios de marca entre consumidores a través del análisis longitudinal de matrices o *brand-switching*.¹⁷

Fuera del ámbito de la Psicología o el Marketing, por ejemplo dentro del campo clínico, el Unfolding de Tres Vías podría aplicarse para examinar el efecto en las preferencias hacia ciertos tratamientos antes y después de su aplicación, e incluso su poder de aplicación puede extenderse hacia otros campos de las ciencias sociales del comportamiento como la Sociología, la Economía, la Política, la Antropología, etc.

A continuación se exponen varios ejemplos de aplicación de los modelos Unfolding de tres vías y de Unfolding de tres vías junto a análisis cluster.

Aplicación del modelo ponderado de tres vías de DeSarbo y Carroll (1985)

Ejemplo 1: DeSarbo and Rao (Pain Reliever Preference Data)

En el artículo de DeSarbo y Carroll (1985), centrado en el Modelo Unfolding de tres vías, de las aplicaciones que realizan los autores del modelo, una de ellas consiste en la aplicación de la técnica sobre un conjunto de datos proveniente de un estudio de DeSarbo y Rao (1983) en el cual se tomó una muestra de 61 estudiantes de un MBA de la Universidad de Fairleigh Dickenson, que puntuaron diez medicamentos contra el dolor (*Anacin, Ascriptin, Aspirina Bayer, Bufferin, Cope, Datril, Excedrin, aspirina Hudson, Tylenol, y Vanquish*) con puntuaciones de 0 a 10 en tres contextos (situaciones) de dolor diferentes (dolor de cabeza, fiebre y dolor muscular). Se

¹⁷Una **Matriz Brand-Switching** es una matriz de dos entradas que indica qué marcas son compradas por una muestra de consumidores en un determinado período y cuáles compran en un período subsiguiente, de esta manera se puede analizar el cambio de preferencia entre marcas así como la fidelidad de los consumidores hacia una determinada marca.

configuró una matriz de tres entradas, donde las filas correspondían a los individuos, las columnas representaban los 10 medicamentos sobre los que se hacía la elección y como tercera vía, los tres contextos de dolor.

Al aplicar el modelo (23) con unos valores $p=2$ en la función de pesos (25) y $D=2$ para el número de dimensiones, se encuentra un modelo final que explica el 51% de la varianza.

La configuración bidimensional de los individuos y las marcas muestra (Fig. 15) que la dimensión 1 parece presentar una clara separación de las marcas con gran publicidad de las no publicitadas o con poco nicho de mercado, mientras que la 2ª dimensión diferencia el formato aspirina (*Ascriptin, Cope, Vanquish, Hudson, Excedrin, Bufferin, Anacin, y Bayer*) de aquellos que no lo son (*Tylenol y Datril*); además, parece que los cuadrantes III y IV contienen la mayor parte de los puntos ideales (individuos) muy cercanos a las marcas más conocidas y publicitadas, ya sean aspirinas o no (es decir, independientemente del contexto de dolor). Los autores concluyen por tanto que en las tres situaciones (dolor de cabeza, fiebre y dolor muscular), la dimensión que más importancia tiene es que el producto tenga un buen nicho de mercado, es decir que sea conocido.

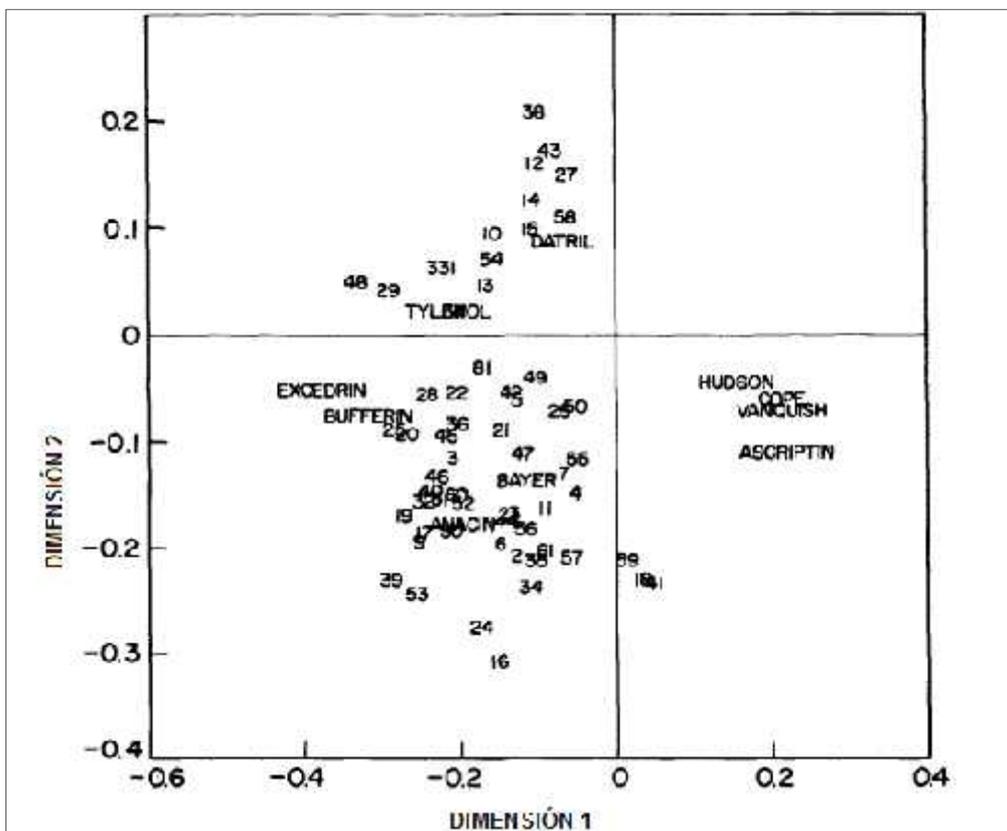


Fig. 15. Representación del espacio conjunto Ejemplo 1 (DeSarbo y Carroll, 1985)

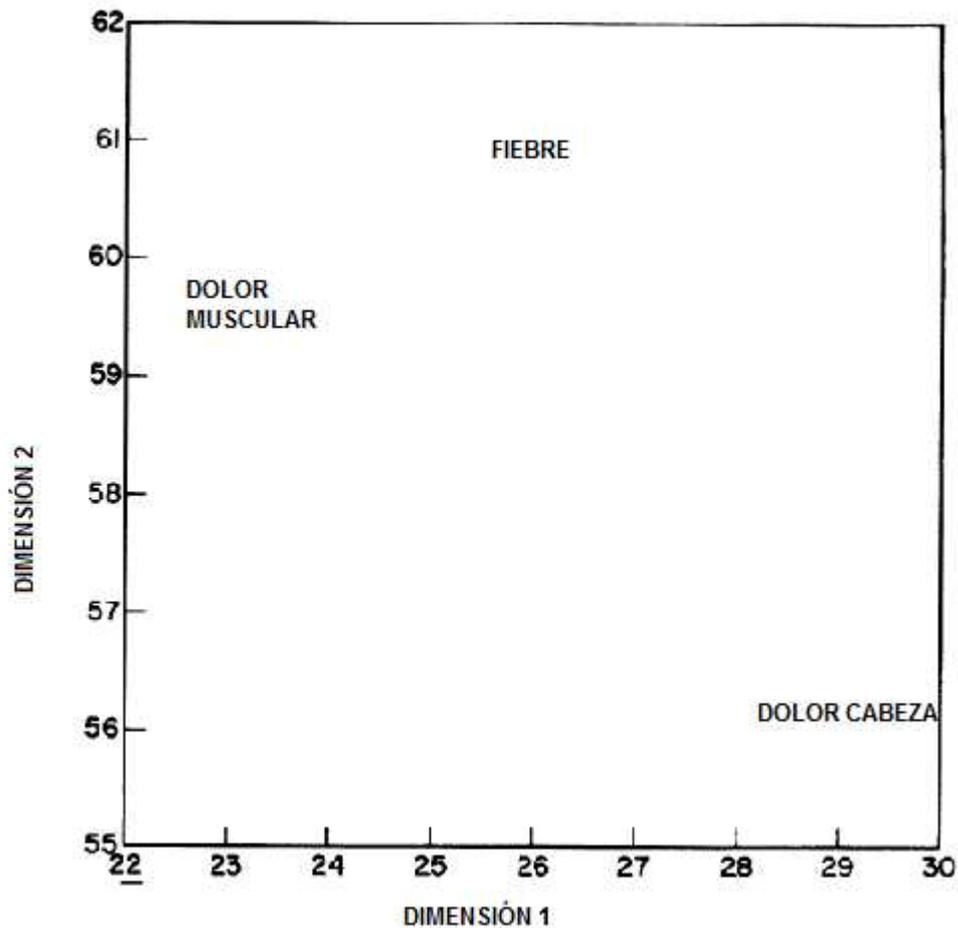


Fig. 16. Pesos obtenidos Ejemplo1 (DeSarbo y Carroll, 1985)

Ejemplo 2: Grambsch, Clark, DeSarbo, and Rothkopf (Reading Profile Data)

Grambsch, Clark, DeSarbo y Rothkopf realizaron en 1983 un experimento para obtener patrones en el movimiento horizontal de los ojos (más conocido como nistagmo) a través de diferentes tareas de lectura. Se pidió a 39 individuos que realizaran seis tareas de lectura diferentes en dos momentos distintos, separados entre sí al menos una semana. En cada una de las ocasiones, los participantes tuvieron que leer cerca de 20.000 palabras, graduadas desde palabras relativamente fáciles hasta textos difíciles, dedicando un tiempo total a la lectura de unos 100 minutos.

Las seis tareas fueron creadas de manera que se diferenciaron entre sí en función del grado de interpretación requerido para su lectura completa y el número de elementos considerados en ella. Su objetivo fundamental fue evaluar cómo el movimiento del ojo (la media de las diferentes medidas) depende de la tarea realizada y averiguar si las diferencias entre tareas pueden o no ser iguales entre individuos.

De las seis tareas de lectura planteadas, cuatro eran de búsqueda y dos de comprensión:

1. La Tarea 1 (*Nonword Search Task*) requería que los individuos buscaran en una lámina un grupo pronunciable de letras sin significado con una o dos sílabas, como por ejemplo hesh. Cuando se encontraba, se reemplazaba por una palabra con contenido en el texto original.
2. La Tarea 2 (*Verbatim Search Task*) requería que los individuos buscaran en una lámina la presencia de una determinada palabra de una o dos sílabas, como por ejemplo sofá.
3. La Tarea 3 (*Synonym Search Task*) requería que los individuos buscaran el sinónimo monosílabo o bisílabo de una palabra también de una o dos sílabas, como por ejemplo little como sinónimo de small.
4. En la Tarea 4 (*Information Task*) los individuos tenían que buscar en el texto la respuesta a una determinada pregunta formulada al inicio.
5. En la Tarea 5 (*Information Memory Task*), similar a la anterior, los individuos tenían que hacer una primera lectura de comprensión y después, con el texto fuera de su vista, se les hacía una pregunta determinada y se les preguntaba si podían contestar tal pregunta a partir de la información dada en el texto.
6. La última tarea (*Learning*) requería la lectura de dos pasajes, uno de ellos de comprensión fácil y el otro más difícil, compuestos por ocho láminas, cada de uno de los cuales contenía 1200 palabras extraídas de la lectura original. El objetivo era obtener una comprensión general de cada uno. Cada uno de los pasajes versaba de un determinado tema y los individuos podían leerlos a su ritmo, de forma que no se pasara a la lámina siguiente si no se había terminado de leer la anterior. Al término del pasaje, los individuos tuvieron que responder un cuestionario de 8 ítems de respuestas cortas.

Utilizaron electrodos situados cerca de los ojos, conectados a ordenadores que registraban y almacenaban los movimientos en los momentos de lectura.

Para determinar las diferencias entre individuos relacionadas con las tareas de lectura se tuvo en cuenta que la lectura se caracteriza por diferentes medidas que dependen unas de otras y se registraron las siguientes:

- i) Tiempo total de lectura (RT), en milisegundos.
- ii) Número de sacudidas (saltitos) de izquierda a derecha (LRS).
- iii) Número de sacudidas de derecha a izquierda (RLS)
- iv) Número de barridos rápidos (*backsweeps*) (BSW).

- v) Duración media de los periodos en los que el ojo no realizó ningún movimiento (FIX).
- vi) Amplitud de las sacudidas de izquierda a derecha (LRA).
- vii) Amplitud de las sacudidas de derecha a izquierda (RLA).

Con toda esta información, se construyó una matriz de 3 modos y 3 vías (sujetos x medidas x tareas) de dimensión (39x7x6), estandarizando las medidas de los estímulos a lo largo de toda la matriz (ya que las unidades de medida eran diferentes entre sí). Tras ciertas pruebas para elegir una dimensión adecuada (D=1 hasta 4) usando el modelo de punto ideal flotante para los individuos, y ciertas pruebas para elegir el exponente p para la construcción de la matriz de pesos (24) (25), se consideró como elección óptima libre de soluciones degeneradas D=1 (1 dimensión) y p=2, con un porcentaje de varianza explicado de 74,4%.

Junto al análisis Unfolding, los autores realizaron además un análisis cluster de la matriz de puntos ideales flotantes con el fin de examinar patrones de lectura similares entre individuos. Mediante el método de la distancia máxima (*complete linkage*) y el correspondiente dendograma, encontraron 3 clusters representados en (Fig. 17) junto a las localizaciones de los estímulos (RT, LRS, RLS, BSW, FIX, LRA y RLA) en cada una de las 6 situaciones.

Si se observa la (Fig. 17), la única dimensión puede ser interpretada como el esfuerzo en la lectura, ya que las medidas de amplitud de los movimientos bruscos o velocidad (LRA y RLA) ocupan un lugar superior o positivo en la dimensión en todas las situaciones, mientras que el resto de estímulos, que miden la fijación de la mirada, se encuentran localizados en la parte inferior o negativa del eje.

Los puntos ideales medios de cada grupo para cada tarea se conectan entre sí a través de líneas discontinuas, pudiendo adivinarse un estilo de lectura entre los grupos basándose en la proximidad entre los puntos ideales medios y la localización de los estímulos.

Por ejemplo, el Grupo 1 está formado por los participantes que son más lentos en la lectura, sobre todo en las tareas 2 y 4, mientras que el Grupo 3 parece ser aquel que contiene a los individuos que leen más rápido y que mejores puntuaciones obtienen en las tareas 3, 4 y 5. El Grupo 2 no muestra evidencias de diferir en cada una de las seis tareas, ya que sus puntuaciones medias son estables en todas ellas.

Los resultados de este análisis de Unfolding de tres vías muestran, por tanto, un ejemplo gráfico de las diferencias existentes en los estilos de lectura de un conjunto de individuos a través de diferentes tareas (que abarcan tanto la búsqueda de palabras y

sinónimos o la comprensión de textos) y, además, examina cómo una serie de aspectos oftalmológicos se ven involucrados en ella.

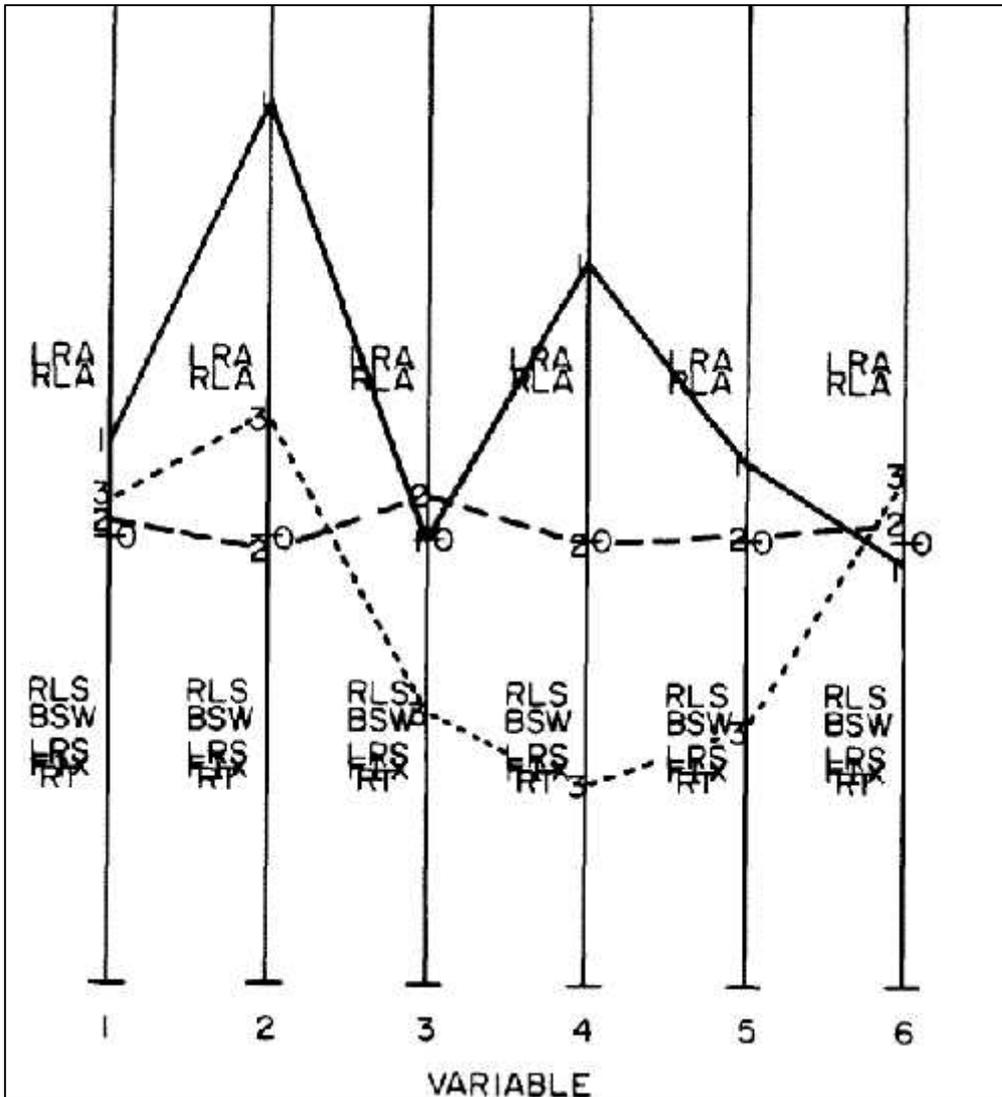


Fig. 17. Espacio conjunto en 1 dimensión Ejemplo 2 (DeSarbo et al., 1985)

Aplicación del Modelo de Tres Vías y Cluster de DeSarbo et al. (2009)

DeSarbo et al. (2009) recurren al estudio de Green y Rao (1972) para aplicar su modelo mixto de Unfolding de tres vías y análisis cluster, introduciendo una tercera vía, la medida de las preferencias en seis contextos o situaciones diferentes:

1. "Preferencia General"
2. "Cuando tomo un desayuno de zumo, bacon, huevos y bebida"
3. "Cuando tomo un desayuno de zumo, cereales y bebida"

4. “Cuando tomo un desayuno de zumo, panecillos, salchicha, y bebida”
5. “Sólo bebida”
6. “A la hora de almorzar solo con bebida”

Añaden además cierta información sociodemográfica: edad, sexo, estado civil y número de hijos, si se consideran con sobrepeso y en cuánto lo estiman y para finalizar cuál es su desayuno típico de la semana y durante el fin de semana, y el consumo habitual de café, té y leche.

En las siguientes tablas, se muestran los códigos de los productos (**Tabla 3**) y los atributos (**Tabla 4**) que se utilizan en las representaciones gráficas.

TP (Toast pop-up): Tostada	BTJ (Buttered toast and jelly): Tostada con mantequilla y gelatina
BT (Buttered toast): Tostada con mantequilla	TMn (Toast and margarine):Tostada y margarina
JD (Jelly donut): Donut relleno	CB (Cinnamonbun): Bollo de canela
CT (Cinnamontost): Tostada con canela	DP (Danish pastry): Pasta danesa
BMM (Blueberrymuffin and margarine): Magdalena de arándanos y margarina	CC (Coffee cake): Pastel de café
HRB (Hard rolls and butter): Panecillos y mermelada	CMB (Corn muffin and butter): Magdalena y mantequilla
TMd(Toast and marmalade): Tostada con mermelada	GD (Glazed Donut): Donut Glacé
EMM (English muffin and margarine): Magdalena inglesa y margarina	

Tabla 3. Productos en ejemplo de Unfolding de Tres Vías y Cluster (Fuente DeSarbo et al. (2009))

R1	1. Facilidad de preparación
R2	2. Sabor
R3	3. Indicado para adultos
R4	4. Calorías
R5	5. Artificial-Natural
R6	6. Textura
R7	7. Coste
R8	8. Sin relleno
R9	9. Momento más adecuado para su consumo
R10	10. Forma
R11	11. Tostado

Tabla 4. Atributos en ejemplo de Unfolding de Tres Vías y Cluster (Fuente DeSarbo et al., 2009)

Tras la aplicación del modelo de Unfolding de tres vías y análisis cluster expuesto en el capítulo anterior a la matriz de tres entradas con los datos arriba indicados, y tras la evaluación de diferentes modelos, los autores consideran una solución óptima en $D=2$ dimensiones y $S=3$ clusters disjuntos, sin pesos ($q_{ijt} = 1$) que, en principio, no parece mostrar soluciones degeneradas.

La representación conjunta de productos y atributos bajo la situación 1 (preferencia general) se muestra a continuación (Fig. 18).

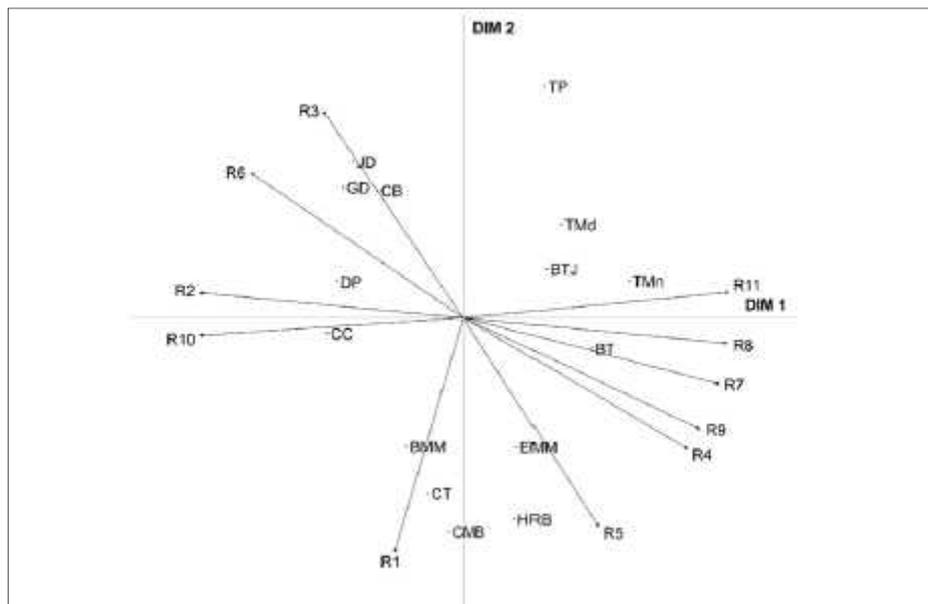


Fig. 18. Espacio conjunto sin clusters (preferencia general) (DeSarbo et al., 2009)

Los tres clusters obtenidos son diferentes entre sí en el sentido de la preferencia, los individuos dentro de un cluster específico se caracterizan por mostrar una preferencia parecida en ese contexto pero, en otro contexto, pueden variar su preferencia y desplazarse de posición respecto de la preferencia general.

El primer cluster (Fig. 19), contiene a los individuos cuya preferencia deriva hacia los productos que contienen algún tipo de tostada (BT: Tostada con mantequilla; TMn: Tostada y margarina) en todas la situaciones ($y_{1,1}, y_{1,2}, y_{1,3}, y_{1,4}, y_{1,5},$) excepto en la situación 6 ($y_{1,6}$) a la hora de almorzar solo con bebida en la que los individuos pertenecientes a ese cluster se decantan por algún tipo de magdalena que es la opción más saludable dentro de los productos dulces. La lectura general puede hacer pensar que este primer cluster realmente agrupa a los individuos cuya preferencia en cualquier tipo de situación, está enfocada a ciertos productos de manera muy directa,

evitando los dulces, los productos calóricos y los rellenos. Se puede deducir que los individuos pertenecientes a este cluster son aquellos preocupados con su salud.

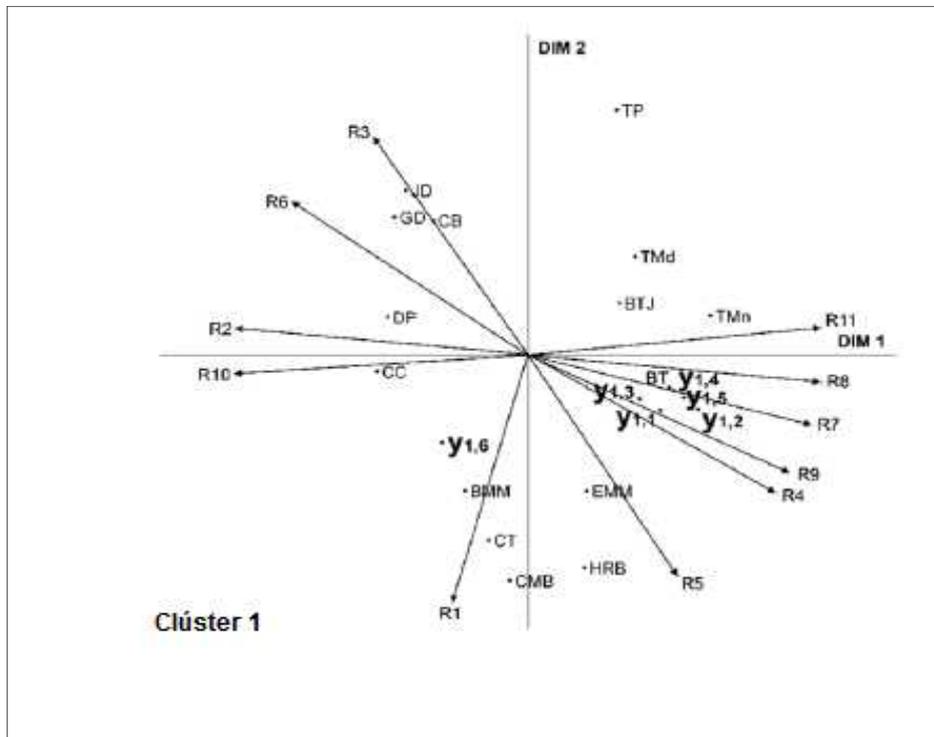


Fig. 19. Espacio conjunto Cluster 1 (DeSarbo et al., 2009)

En el caso del Cluster 2 (Fig. 20), se puede observar que las preferencias cambian claramente en comparación con el cluster 1. Por un lado, estos individuos sí que muestran preferencia hacia determinados productos en situaciones distintas: en las situaciones de preferencia general, a la hora de la bebida y en el almuerzo (y_{21}, y_{25}, y_{26}) los individuos pertenecientes a este cluster muestran una tendencia hacia los dulces y productos altos en calorías como los pasteles de café (CC) y las pastas danesas (DP), mientras que en las otras tres situaciones (y_{22}, y_{23}, y_{24}) prefieren opciones más simples y saludables como las tostadas (TP). Esto puede llevar a pensar que estos individuos tienen una tendencia a llevar un cierto equilibrio en su dieta ya que sólo consumen productos calóricos cuando los toman con bebida o bien los toman en el almuerzo, mientras que optan por la tostada cuando desayunan en compañía de más productos como el zumo, bacon, huevos o salchichas.

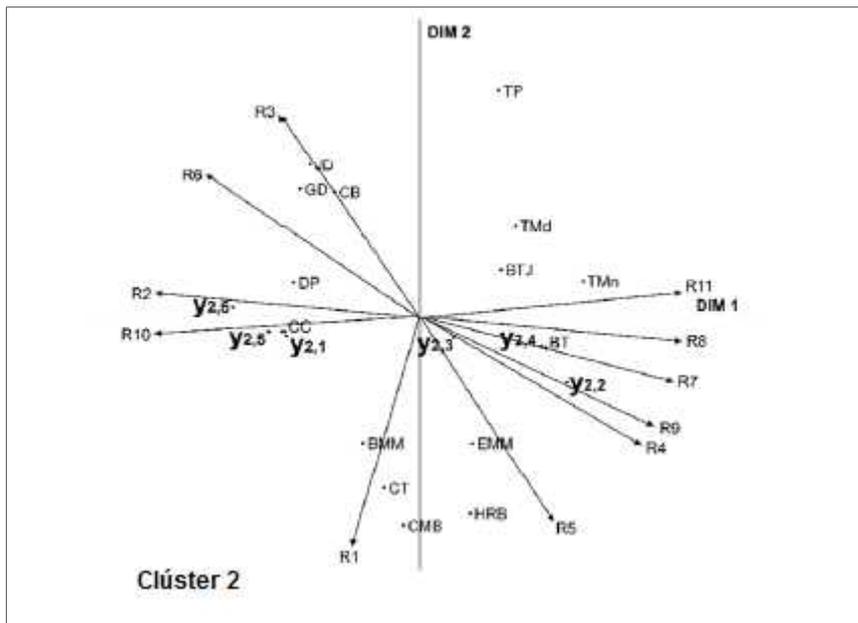


Fig. 20. Espacio conjunto Cluster 2 (DeSarbo et al., 2009)

Para finalizar el Cluster 3 (Fig. 21), agrupa a los individuos cuya preferencia en casi todas las circunstancias ($y_{3,1}, y_{3,3}, y_{3,4}, y_{3,5}, y_{3,6}$) apunta hacia los dulces y los productos altos en calorías y que no parecen cambiar de gustos sea cual sea el contexto, excepto en el caso de desayuno con zumo, bacon, huevos y bebida ($y_{3,2}$) en el que optan por productos tipo tostada. En este caso, el cluster puede agrupar individuos que preferiblemente eligen en casi todas las ocasiones productos dulces y con calorías.

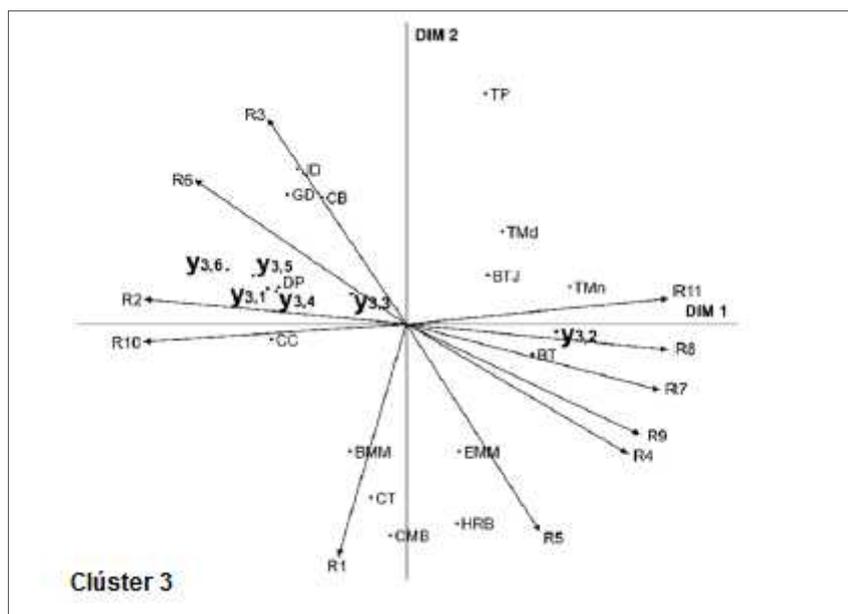


Fig. 21. Espacio conjunto Cluster 3 (DeSarbo et al., 2009)

Además, la matriz de pertenencia P proporciona sustanciosa información, ya que en ella se conoce qué individuo de la muestra pertenece a cada cluster, y su combinación con la información sociodemográfica puede caracterizar mejor los clusters que el modelo ha estimado. Por ejemplo en este caso, tras este cruce de información, los autores comentan que el cluster 1, formado por individuos que se preocupan de su salud además es en su mayoría de sexo masculino, de edad avanzada, con hijos y que beben café; el cluster 2, formado por aquellos individuos que llevan una dieta equilibrada es mayoritariamente femenino, soltero, sin sobrepeso, que beben té y tienden hacia la comida con pocas calorías; y el cluster 3, formado por individuos que consumen dulces y calorías está compuesto de individuos de sexo masculino, de edad joven, casados, con sobrepeso y que prefieren beber café con leche y desayunos contundentes.

En general los clusters 1 y 3 tienen preferencias estables en la mayoría de las situaciones, mientras que para los individuos del cluster 2 las preferencias cambian según el contexto.

Implementación del Algoritmo Weighted Least Squares Algorithm for Three-Way Unfolding de 1985 en Código MATLAB

En esta sección se describe una aportación práctica del Algoritmo para Unfolding de 3 Vías explicado en el capítulo anterior, que consiste en el desarrollo del programa **Algoritmo_3WU** implementado en Matlab© R2012.

Tras la revisión realizada en este trabajo del Software existente para realizar análisis MDU, es relevante la escasa disponibilidad de software para realizar Unfolding de 3 Vías exceptuando a PREFSCAL, que brinda al usuario la opción de análisis de preferencias para 2 ó 3 vías. Ese ha sido principalmente el aliciente que ha motivado el poder desarrollar una herramienta que calcule soluciones Unfolding para datos de 2 o 3 vías y en un futuro tener la posibilidad de mejorarla y aportarle nuevas funcionalidades, ya que el algoritmo de DeSarbo et al. (1985) no se encuentra disponible en ningún software conocido ni desarrollado en ningún lenguaje de programación al menos de manera pública.

Se ha considerado la programación de este algoritmo en Matlab debido a las características de este software, orientado principalmente al cálculo matricial, optimización y simulación de procesos. No obstante, sería también factible su implementación en R, ya que DeLeeuw y Mair (2009) han desarrollado el paquete

Smacof (tanto para MDS como para Unfolding), e incluso en FORTRAN, pues existen numerosas rutinas de algoritmos que calculan soluciones MDS, todas ellas disponibles en el siguiente enlace <http://www.netlib.org/mds/index.html> (por ejemplo, PREFSCAL o MDPREF), desarrolladas en este lenguaje de programación orientado al cálculo numérico y la computación científica y cuyos orígenes se remontan a mediados de la década de los 50.

El programa propuesto, Algoritmo_3WU, se compone de un programa principal y 16 funciones (subrutinas) que vienen declaradas al inicio y que en el transcurso del programa son utilizadas con diferentes parámetros de entrada y salida.

A partir de una configuración matricial inicial aleatoria de individuos, estímulos y pesos, realiza por un lado la representación conjunta de las coordenadas óptimas para individuos y estímulos en un espacio de dimensión $D=2$ una vez normalizadas, y por otro lado representa las componentes de matriz de pesos W o importancia para cada situación, y la representación de los individuos en cada situación (lo que DeSarbo llama puntos flotantes) de manera que la función de pérdida (diferencia cuadrada entre las no-preferencias iniciales y las estimadas mediante el algoritmo) alcance un valor mínimo tal y como describen DeSarbo y Carroll (1985).

También se brinda al usuario la opción de utilizar datos de no-preferencia (inversamente relacionados con los de preferencia) (en texto o Excel) obtenidos en un juego de preferencias, simplemente cargando el fichero de texto u hoja Excel.

Para comprobar la funcionalidad del programa, se ha procedido a simular varios escenarios con distintos valores de tamaño en el número de individuos, estímulos y situaciones, manteniendo fijo el número de dimensiones $D=2$. Siguiendo a DeSarbo (1985), no se han ejecutado pruebas de más de 20 individuos debido al elevado coste computacional del proceso.

Las pruebas se han llevado a cabo en un ordenador portátil con Sistema Operativo Windows 7 Home Premium 64 bits, Intel® Core™i5-3317U CPU @ 1,70 GHz y memoria RAM 4 GB. En la **(Tabla 5)** se muestra los resultados de las simulaciones ejecutadas, la tolerancia (TOL), el número de iteraciones necesarias para la convergencia (ITER), el VAF (*Variance Accounted For*) entre las no-preferencias originales y las calculadas y el tiempo computacional (TC) en segundos, como medida total del proceso. En general conforme la muestra aumenta se hace más costoso desde el punto de vista computacional, así que quizás con una máquina más potente (procesador y más memoria), se podrían conseguir mejores resultados y elevar el tamaño muestral.

Simulación	I :Individuos	J :objetos	T :situaciones	p	TOL	ITER	VAF	TC (sg)
1	3	2	2	2	10e(-2)	11	0,9959	1.020,55
2	4	3	2	3	10e(-1)	31	0,9857	3.973,71
3	4	3	2	5	10e(-1)	31	0,9985	1.661,36
4	8	4	3	2	10e(-1)	31	0,9974	54.930,51
5	15	5	2	3	10e(-1)	21	0,9984	3.648,21
6	20	7	3	2	10e(-1)	31	0,9996	21.882,58

Tabla 5. Tabla de resultados de simulaciones Algoritmo 3 Vías DeSarbo (1985)

A continuación se muestra un ejemplo de la representación conjunta que da el programa, en el caso de la representación de los individuos a través del punto fijo (**Fig. 22**) o bien en el caso de que los individuos se representen mediante puntos flotantes (según las situaciones) (**Fig. 23**).

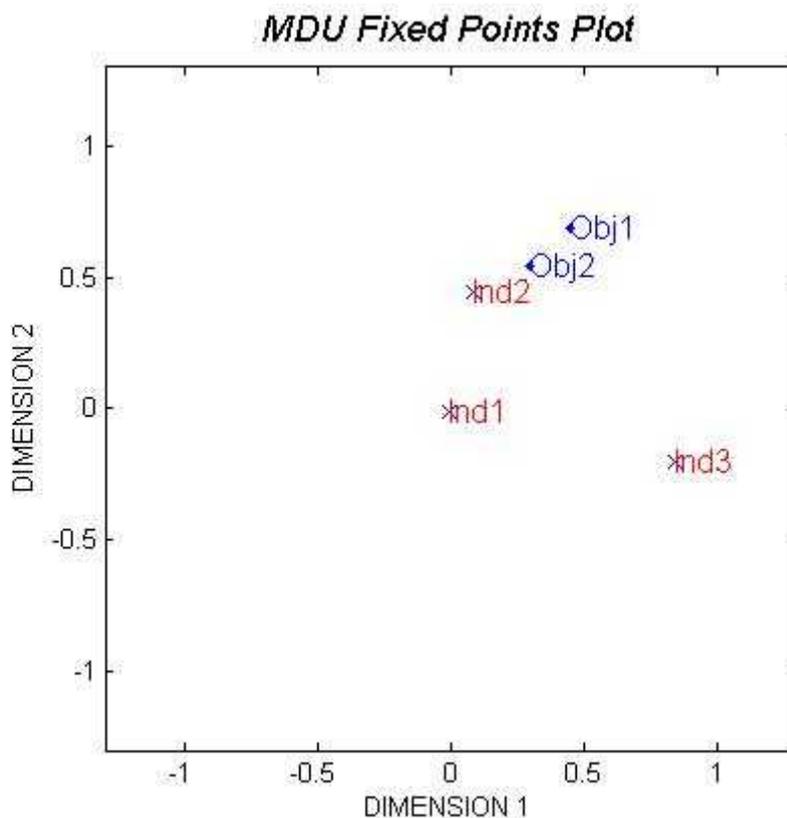


Fig. 22. Ejemplo de Representación Conjunta con puntos fijos en la simulación Matlab Algoritmo 3 Vías DeSarbo (1985)

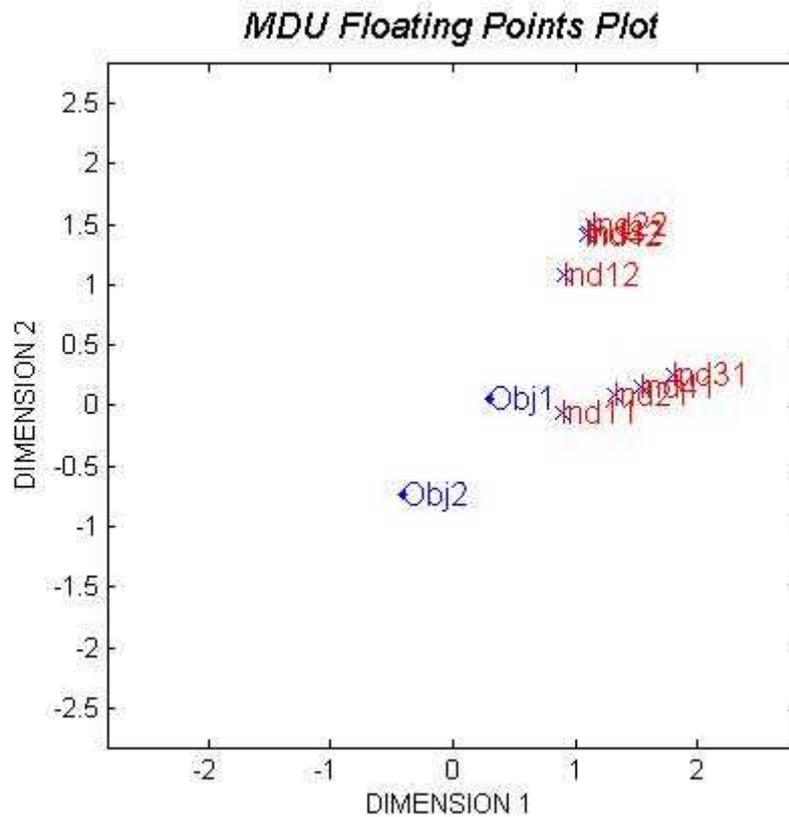


Fig. 23. Ejemplo de Representación Conjunta con puntos flotantes en la simulación Matlab Algoritmo 3 Vías DeSarbo (1985)

No obstante, el programa propuesto podría ser modificado en un futuro mediante una interfaz gráfica que facilite la interacción del programa con el usuario (a través de la interfaz gráfica GUI de Matlab), a fin de poder manipular los gráficos de salida y dar elección al usuario de cambiar colores, símbolos, etc. También cabe la posibilidad de la extensión del algoritmo a datos no métricos o incluso la opción de Unfolding de 3 Vías y Análisis Cluster.


```

Xinic=randn(J,D);

%La matriz Y de los puntos de los individuos I en la nueva
%configuración de dimensión IxD a partir de una distribución N(0,1)

Yinic=randn(I,D);
erroral=randn(T,I*J)/100;

%Las variables a estimar en forma simbólica que es la incógnita a
%despejar

Xreal=sym('x',[J,D]); %Estímulos
Yreal=sym('x',[I,D]); %Individuos

%_____0: Configuración Inicial _____

%Este paso es solo válido para ejecutar las simulaciones

X=Xinic;
Y=Yinic;
W=Winic;
alpha=alphainic;
[Deltainic Gammainic]=Configinicial(X,Y,W,alpha,erroral,p);

%_____1: Datos de Despreferencias _____

%Si se va a importar directamente datos de despreferencias usar un
%comando para importar ficheros de texto:load o textread o
%excel:xlsread
%fichero :ruta y nombre fichero

%Deltainic=load(fichero.txt);
%Deltainic=textread('fichero.txt');
%Deltainic= xlsread('fichero.xls');

%Calcular a continuación los pesos Gamma
%Gammainic=1./Deltainic.^p;

%_____2:Estimar los Estímulos X _____

%Se retienen fijos W,Y y alpha dejando libres los valores de X que
%es sobre lo que se va a optimizar

[Z2 DZ2x]=Fobjetivo(Y,Xreal,W,alpha,Gammainic,Deltainic,1);
X0=randn(J,D); %Punto inicial
[xup f]= Grad_Conjugados_unc(Z2,DZ2x,Xreal,X0,1,100);
Xup=VectoraMatriz(xup);

%_____3:Estimar los Individuos Y _____

%Se retienen fijos W,X y alpha dejando libres los valores de Y que es
%sobre lo que se va a optimizar

[Z2 DZ2y]=Fobjetivo(Yreal,X,W,alpha,Gammainic,Deltainic,2);
Y0=randn(I,D);%Punto inicial
[yup f] = Grad_Conjugados_unc(Z2,DZ2y,Yreal,Y0,1,100);
Yup=VectoraMatriz(yup);

```

```

%_____ 4:Estimar W y alpha _____

[Wup,alphaup] = EstimaParametros(Deltainic,Yup,Xup);

%_____ 5:Criterio de parada /Tolerancia _____

[zold zzold]=Fobjetivo(Y,X,W,alpha,Gammainic,Deltainic,0);
[znow zznw]=Fobjetivo(Yup,Xup,Wup,alphaup,Gammainic,Deltainic,0);

% Se itera hasta que la diferencia entre lo estimado y el valor
inicial es mínima:se alcanza la convergencia
dif = realmax;

%Se inicializa el contador
numiter = 1;
dif =abs(zold-znow);
zold=znw;
X=Xup;
Y=Yup;
W=Wup;
alpha=alphaup;
%_____ Bucle Central _____

warning off
while dif> tol & numiter <= total
numiter = numiter + 1;
%Actualiza X:estímulos
[Z2 DZ2x]=Fobjetivo(Y,Xreal,W,alpha,Gammainic,Deltainic,1);
X0=rand(J,D);
[xup f] = Grad_Conjugados_unc(Z2,DZ2x,Xreal,X0,1,100);
Xup=VectoraMatriz(xup);
%Actualiza Y:individuos
[Z2 DZ2y]=Fobjetivo(Yreal,X,W,alpha,Gammainic,Deltainic,2);
Y0=rand(I,D);
[yup f] = Grad_Conjugados_unc(Z2,DZ2y,Yreal,Y0,1,100);
Yup=VectoraMatriz(yup);
clear c
%Actualiza W y alpha
[Wup,alphaup] = EstimaParametros(Deltainic,Yup,Xup);
[znow zznw]=Fobjetivo(Yup,Xup,Wup,alphaup,Gammainic,Deltainic,0);
dif=abs(zold-znow);
fprintf('%4.2f',dif);
zold=znw;
X=Xup;
Y=Yup;
W=Wup;
alpha=alphaup;
clear X0 Y0;
end;

%_____ 6:Normalización _____

[Xcoord Ycoord Wnor]=Normaliza(X,Y,W);

%_____ Ajuste:VAF _____

VAF= CalculaVAF(Ycoord,Xcoord,Wnor,alpha,Gammainic,Deltainic );

%_____ 7:Puntos Flotantes para individuos _____

```

```

YF = Puntos_flotantes(Y,X,Wnor,Deltainic,Gammainic);

%Para el reloj para contar el tiempo de ejecución
toc;
%_____8: Representación gráfica _____

[inclabel objlabel wlabel yflabel] = Labels(Ycoord,Xcoord,Wnor);

%Individuos y Estímulos

v=MDUplot(Xcoord,Ycoord,objlabel,inclabel,1);

%Situaciones

w=Wplot(Wnor,wlabel);

%Puntos flotantes (situaciones) para individuos

f=MDUplot(Xcoord,YF,objlabel,yflabel,2);

```

A continuación se declaran las funciones necesarias para la ejecución:

```

function [C1 C2]=Distancia(A,B)
%Función que calcula la distancia entre todos los puntos de ambas
matrices

[r d]=size(A);%Individuos
[n d]=size(B);%Estímulos
for i=1:r
    for j=1:n
        C1(i,j)=(A(i,1)-B(j,1))^2;
        C2(i,j)=(A(i,2)-B(j,2))^2;
    end;
end;

function [Delta Gamma]=Configinicial(X,Y,W,alpha,erroral,p)
%Función que calcula el Delta y Gamma inicial de manera aleatoria

[r d]=size(Y);
[n d]=size(X);
[t d]=size(W);
[C1 C2]=Distancia(Y,X);
D=[reshape(C1.',1,[]); reshape(C2.',1,[])] ;
D2=W*D;
clear C1C2
for t=1:t
    Deltint(t,:)=D2(t,:)+alpha(t,:)
    Delta(t,:)=Deltint(t,:)+erroral(t,:);
end;
clear Deltint;
%La matriz gamma(t,i,j)
Gamma=1./Delta.^p;

function der=Deriva(f,D)
%Esta función calcula las derivadas parciales de la función objetivo
%Z2 respecto de todas las variables de X(d_xjd) y respecto de Y(d_yid)

[n m]=size(D);

```

```

for ind=1:n
for d=1:m
    der(ind,d)=diff(f,D(ind,d));
end
end;

function [z g]= Fobjetivo(Y,X,W,alpha,gamma,Deltafin,par)
%Esta función calcula la función objetivo a minimizar z y su gradiente
%respecto de las variables X (par=1) o Y (par=2): g

[r d]=size(Y);
[n d]=size(X);
[t d]=size(W);

[C1 C2]=Distancia(Y,X);
D=[reshape(C1.',1,[]); reshape(C2.',1,[])] ;
D2=W*D;
for t=1:t
%Para la situación t=1,2..T
    Delta(t,:)=D2(t,:)+alpha(t,:);
end;
for i=1:t
    for k=1:r*n
        z2(i,k)=gamma(i,k).*(Deltafin(i,k)-Delta(i,k)).^2;
    end;
end;
z=sum(sum(z2)); %funcion objetivo;
if par==1
    der=Deriva(z,X);
    g=reshape(der.',1,[]).';
elseif par==2
    der=Deriva(z,Y);
    g=reshape(der.',1,[]).';
else %cualquier otro valor de par
display ('En este caso se obtiene sólo el valor de la función objetivo
evaluada');
for i=1:t
    for k=1:r*n
        z2(i,k)=gamma(i,k).*(Deltafin(i,k)-Delta(i,k)).^2;
    end;
end;
g=0;
end;

function [z g V v0] = Sustituye(f,g,X,X0)
%Función que sustituye un valor X0 en la función objetivo y el
%gradiente, z es la función objetivo, grad es el gradiente, X es la
%matriz de parámetros desconocidos y X0 es el valor correspondiente

V=reshape(X.',1,[]).';
v0=reshape(X0.',1,[]).';
z=subs(f,V,v0);
g=subs(g,V,v0);
end

function [x f] = Grad_Conjugados_unc(obj,der,Xreal,X,par,itera)
% Método de los gradientes conjugados para minimizar f(x)

rho = 0.01; beta = 0.1; % Parámetros de line search

```

```

maxit = itera;
[f g U x]=Sustituye(obj,der,Xreal,X);
%[f g]= fun(x);
ppr=-g; gpr=g; pbeta=0; % Comienzo con máx. pendiente
for i=1:maxit % Proceso iterativo
if i>1
[f g]=Sustituye(obj,der,Xreal,X);
if par==1, pbeta=(g'*g)/(gpr'*gpr); % Fletcher-Reeves
else pbeta=((g-gpr)'*g)/(gpr'*gpr); % Polak-Rivière
end
end
p = -g+pbeta*ppr;
if (norm(g,inf) < 1e-6), break, end% Condición de mínimo
alpha = 1;
for k=1:10 % Backtracking amplitud de paso
xnew = x+alpha*p; fxnew = Sustituye(obj,der,Xreal,xnew);
if fxnew <= f + alpha*rho*g'*p
break
else alpha = alpha*beta;
end
end
x = x + alpha*p; ppr=p; gpr=g;
fprintf( '%4.0f %13.8e %13.8e %13.8e %13.8e\n',i,x,f,alpha);
end

function M = VectoraMatriz(v)
%Esta función convierte un vector v en una matriz M

d=2;
M=vec2mat(v,d);
end

function [Wup alphaup] = EstimaParametros(Deltainic,Y,X)

%Función calcula la matriz W de constantes multiplicativas y
%el vector de constantes aditivas alpha

[m d]=size(Y);
[n d]=size(X);
[C1,C2]=Distancia(Y,X);
p1=reshape(C1.',1,[]);
p2=reshape(C2.',1,[]);
unoP=ones(m*n,1);
P=[unoP p1 p2];
Parest=Deltainic*P*inv(P'*P);
alphaup=Parest(:,1);
Wup=Parest(:,2:end);
end

function [Xc Yc Wnor] = Normaliza(X,Y,W)
%Función que Normaliza las matrices de Individuos, Estímulos y
Constantes

[r d]=size(Y);
[n d]=size(X);
J=[X;Y];
N=normc(J);%Normaliza por columnas
[j1 j2]=size(J);
for i=1:j2
c(i)=sqrt(sumsqr(J(:,i)));
end;

```

```

if j2==d
for i=1:j2
    Wnor(:,i)=c(i)*W(:,i);
end;
else
    display ('Error: el número de columnas de ambas matrices debe ser
igual');
end;
Xc=N(1:n,:);
Yc=N(n+1:end,:);

```

```

function VAF= CalculaVAF(Y,X,W,alpha,gamma,Deltafin)
%Esta función calcula el VAF entre las no-preferencias originales y
%las estimadas en el nuevo espacio de dimensión reducida D=2

```

```

[r d]=size(Y);
[n d]=size(X);
[t d]=size(W);
[C1 C2]=Distancia(Y,X);
D=[reshape(C1.',1,[]); reshape(C2.',1,[])] ;
D2=W*D;
for t=1:t
%Para la situación t=1,2..T
    Delta(t,:)=D2(t,:)+alpha(t,:);
end;
for i=1:t
    for k=1:r*n
        z2(i,k)=gamma(i,k).*(Deltafin(i,k)-Delta(i,k)).^2;
    end;
end;
z=sum(sum(z2));
m=sum(Delta,2)/r*n;
for i=1:t
    for k=1:r*n
        den(i,k)=gamma(i,k).*(((Deltafin(i,k))-m(i)).^2);
    end;
end;
zd=sum(sum(den));
VAF=1-(z/zd);
end

```

```

function YF = Puntos_flotantes(Y,X,W,Deltainic,Gammainic)
%Función que calcula los s*r puntos Y flotantes

```

```

[r d]=size(Y);
[n d]=size(X);
[s d]=size(W);
b=reshape((Deltainic.*Gammainic).',n,[]).';
D=[]; %Se almacena una matriz de Coeficientes
P=[]; %Se almacena la matriz de parámetros estimados
F=[];
for t=1:s %número de situaciones
    for j=1:n %número de atributos o estímulos
%por cada t:situación y para todos los i:individuos , siempre
%hay 3 parámetros a estimar
        C(j,1)=W(t,1)*X(j,1).^2+W(t,2)*X(j,2).^2;
        C(j,2)=W(t,1)*X(j,1);
        C(j,3)=W(t,2)*X(j,2);
    end;
end;

```

```

        D=[D;C];
        l=(t*r-(r-1));
        m=t*r;
        clear p;
for i=1:m %número de individuos
    p(:,i)=pinv(C)*b(i,:).';
    P=[P,p(:,i)];
    f(1,i)=p(2,i)/p(1,i);
    f(2,i)=p(3,i)/p(1,i);
    F=[F,f(:,i)];
end;

        clear p;
        clear C;
end;
YF=F.';
end

function [indlabel objlabel wlabel yflabel] = Labels(Y,X,W)
%Función que genera automáticamente etiquetas para los Individuos,
Objetos, Constantes y Puntos Flotantes

[r d]=size(Y);
[n d]=size(X);
[t d]=size(W);
ind = cell(1,r);
for i=1:r
    ind{i} = ['Ind' num2str(i)];
end
indlabel=char(ind);
obj= cell(1,n);
for i=1:n
    obj{i} = ['Obj' num2str(i)];
end
objlabel=char(obj);
weight = cell(1,t);
for i=1:t
    weight{i} = ['W' num2str(i)];
end
wlabel=char(weight);
floating = cell(r,t);
for j=1:r
for i=1:t
    floating{j,i} = ['Ind' num2str(j) num2str(i) ];
end
end
yflabel=char(floating);

function p=MDUplot(X,Y,columnlabel,rowlabel,par)
%Función que representa de manera conjunta los individuos Y y los
objetos X

[n d]=size(X);
r=size(Y,1);
plot(Y(:,1),Y(:,2),'x','MarkerSize',8);
if par==1
title('\it{MDU Fixed Points Plot}','FontSize',14);
elseif par==2
title('\it{MDU Floating Points Plot}','FontSize',14);
end
end

```

```

xlabel('DIMENSION 1');
ylabel('DIMENSION 2');
hold on
norm=max(EuclD([Y;X],zeros(1,d)));
axis square
axis equal
%set(gca,'XTick',[],'YTick',[])
axis([-1.5*norm 1.5*norm -1.5*norm 1.5*norm])
if size(columnlabel,1)==n
    plot(X(:,1),X(:,2),'.','MarkerSize',8);

text(X(:,1),X(:,2),columnlabel,'FontSize',12,'FontWeight','normal','color','b' );
end;
if size(rowlabel,1)==r
    text(Y(:,1),Y(:,2),rowlabel,'FontSize',12,'color','r');
end;
hold off

function w = Wplot (X,Xlabels)
%Función que representa las constantes multiplicativas de cada
%situación en la misma escala que X e Y
[n d]=size(X);
title('\it{Weight Plot}','FontSize',14)
xlabel('DIMENSION 1');
ylabel('DIMENSION 2');
hold on
axis square
axis equal
%set(gca,'XTick',[],'YTick',[])
if size(Xlabels,1)==n
    plot(X(:,1),X(:,2),'.','MarkerSize',8);

text(X(:,1),X(:,2),Xlabels,'FontSize',12,'FontWeight','normal','color','b' );
end;
hold off
end

```

Conclusiones

Tras exponer de manera rigurosa las características del modelo Unfolding y su desarrollo hasta el momento actual, se puede concluir que:

1. Se ha realizado una revisión conceptual de la técnica, abordándola a través del Álgebra Lineal (matrices de disimilaridades/no-preferencias y transformaciones) y la Geometría (nociones y conceptos de distancias y espacios euclídeos presentes a lo largo de todo el trabajo).
2. Se han recopilado y expuesto una serie de pasos y aproximaciones al problema de las soluciones degeneradas y triviales. Si bien alguna de éstas no ha dado un buen resultado, se ha llegado a mostrar que la técnica propuesta por Busing (2005) del coeficiente de penalización funciona y está integrada hoy en día en el software comercial disponible para los usuarios.
3. Se han expuesto los diferentes modelos y tipos de Unfolding existentes, así como sus expresiones matemáticas, en base al desarrollo matricial y geométrico de cada modelo y el tipo de configuraciones que generan.
4. Se ha revisado cronológicamente el software disponible para realizar análisis Unfolding, citando los programas, autores, disponibilidad en la actualidad, tipo de modelos que analizan y otro tipo de características.
5. Tal y como ocurre en otras disciplinas del AM, se ha introducido el concepto de tabla de tres entradas, así como la propuesta de DeSarbo (1985) a través del modelo ponderado de mínimos cuadrados alternos para encontrar soluciones a datos de preferencias de tres entradas.
6. Como apunte importante se ha revisado una de las últimas aportaciones realizada por DeSarbo et al. (2009) que consiste en la fusión del Unfolding de tres vías junto al análisis cluster, como forma no sólo de representación conjunta de individuos y estímulos en un espacio común en orden a preferencias establecidas y diferentes contextos, sino como forma de agrupar individuos en base a ciertas dimensiones no percibidas en sus preferencias a lo largo de diferentes situaciones.

7. Para finalizar se han expuesto diversas aplicaciones de Unfolding, Unfolding de tres vías y Unfolding de tres vías y análisis cluster a diversos campos, fundamentalmente en Psicología y Marketing (Análisis de Mercado) en los que está más extendido, aunque se han encontrado también aplicaciones en Sociología, Política, Pedagogía, Psicología Clínica y Genética. Ciertamente su uso es posible en cualquier disciplina en el que un conjunto de individuos participen en un proceso de elección de cualquier clase de objeto o estímulo, en un determinado momento o en una serie de situaciones distintas ya sean temporales o de otras características.

8. Como aplicación práctica se ha desarrollado en Software (Matlab) uno de los algoritmos propuestos en este trabajo y se han ejecutado diversas simulaciones en función del número de individuos, atributos y situaciones para contrastar su funcionamiento.

Bibliografía

- Addelman, S. (1962). Orthogonal Main Effects Plans for Asymmetrical Factorial Experiments. *Technometrics*, 4, 21-46.
- Balabanis, G., Diamatopoulos, A. (2004). Domestic country bias, country of origin effects and consumer ethnocentrism: a multidimensional unfolding approach. *Academy of Marketing Science Journal*, 32 (1), 80-85.
- Belk, R.W. (1974). An exploratory assessment of situational effects in buyer behavior. *Journal of Marketing Research*, 11(2), 156–163.
- Belk, R.W. (1975). Situational variables and consumer behavior. *Journal of Consumer Research*, 2(3), 157–164.
- Bennett, J. F., & Hays, W. L. (1960). Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika*, 25, 27–43.
- Böckenholt, U., Böckenholt, I. (1991). Constrained latent class analysis: Simultaneous classification and scaling of discrete choice data. *Psychometrika*, 56 (4), 699–716.
- Borg, I., Groenen, P. (2005). *Modern Multidimensional Scaling*. New York: Springer.
- Busing, F. M. T. A., Groenen, P. J. F., Heiser, W. J. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, 70(1), 71–98.
- Busing, F. M. T. A., Heiser, W. J., Neufeglise, P., Meulman, J. J. (2005). PREFSCAL. Program for metric and nonmetric multidimensional unfolding, including individual differences modeling and fixed coordinates. SPSSInc. Chicago, IL. (Version 14.0)
- Busing, F. (2010). *Advances in Multidimensional Unfolding*. Doctoral thesis, Leiden University, The Netherlands.
- Carroll, J. D. (1972). Individual Differences and Multidimensional Scaling. In R.N. Shepard, A. K. Romney & S. Nerlove (Eds.), *Multidimensional Scaling: Theory and Applications in the Behavior Sciences, Vol. I: Theory*. New York: Seminar Press.
- Carroll, J. D. (1980). Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E. Lantermann & H. Feger (Eds.), *Similarity and choice*. Bern, Switzerland: Huber.
- Carroll, J. D., Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, 35, 283–319.
- Carroll, J.D., Chaturvedi, A. (1995). A general approach to clustering and multidimensional scaling of two-way, three-way, or higher data. In R. D.Luce et al. (Eds.), *Geometrical Representations of Perceptual Phenomena*. Mahwah, NJ: Lawrence Erlbaum.
- Carroll, J.D., Chang, J.J. (1969). How to use MDPREF, a Computer Program for Multidimensional Analysis of Preference Data. Bell Laboratories.
- Chatterjee, R., DeSarbo, W. (1992). Accommodating the effects of brand unfamiliarity in the multidimensional scaling of preference data. *Marketing Letters*, 3(1), 85-99.

- Chessel, D., Hanafi, M. (1996). Analyses de la Co-Inertie de K nuages de points. *Revue de Statistique Appliquée*, 44(2), 35- 60.
- Chintagunta, P.K. (1994). Heterogeneous logit implications for brand positioning. *Journal of Marketing Research*, 31, 304–311.
- Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 148–158.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Coombs, C.H, Avrunin, G.S. (1977). Single peaked functions and the theory of preference. *Psychological Review*, 84, 216-230.
- Coombs, C.H, Avrunin, G.S. (1988). *The Structure of Conflict*. Hillsdale, NJ: Erlbaum.
- Cox, T.F, Cox, M.A. (2000). *Multidimensional Scaling*. New York: Chapman & Hall/CRC.
- Cuadras, C.M. et al. (1985). Métodos Geométricos en la Estadística. *Queestió*, 9(4), 219-250.
- Cuadras, C. M. (2012). *Nuevos Métodos de Análisis Multivariante*. Barcelona: CMC Editions.
- De Leeuw, J. (1983). On degenerate nonmetric unfolding solutions (Tech. Rep.). Leiden, The Netherlands: Department of Data Theory, FSW/RUL.
- De Leeuw, J. (2005). Multidimensional Unfolding. In *Encyclopedia of Statistics in Behavioral Science* (pp. 1289-1294). Chichester, West Sussex: John Wiley & Sons.
- De Leeuw, J., Heiser, W. J. (1977). Convergence of correction-matrix algorithms for multidimensional scaling. In J. C. Lingoes, E. E. C. I. Roskam, & I. Borg (Eds.), *Geometric representations of relational data* (pp. 735–752). Ann Arbor, MI: Mathesis Press.
- De Leeuw, J., Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3), 1-30.
- De Rooij, M., Heiser, W.J. (2000). Triadic distance models for the analysis of asymmetric three-way proximity data. *British Journal of Mathematical and Statistical Psychology*, 53, 99–119.
- DeSarbo, W. S. (1978). *Three-way unfolding and situational dependence in consumer preference analysis*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.
- DeSarbo, W. S., & Rao, V. R. (1983). *A constrained unfolding model for product positioning*. Unpublished manuscript, Bell Laboratories, Murray Hill, NJ.
- DeSarbo, W. S., Rao, V.R. (1984). GENFOLD2: A set of models and algorithms for the GENERALUnFOLDing analysis of preference/dominance data. *Journal of Classification*, 1, 147-186.
- DeSarbo, W. S, Carroll, J.D. (1985). Three way metric unfolding via alternating weighted least squares. *Psykometrika*, 50(3), 275-300.

- DeSarbo, W., DeSoete, G., Eliashberg, J. (1987). A new stochastic multidimensional unfolding model for the investigation on paired comparison consumer preference/choice data. *Journal of Economic Psychology*, 8, 357-384.
- DeSarbo, W.S., Howard, D.J., Jedidi, K. (1990). MULTICLUS: A new method for simultaneously performing multidimensional scaling and cluster analysis. *Psychometrika*, 56 (1), 121–136.
- DeSarbo, W.S., Jedidi, K.J., Cool, O. (1991). Simultaneous multidimensional unfolding and cluster analysis: An investigation of strategic groups. *Marketing Letters*, 2 (2), 129–146.
- DeSarbo, W.S., Manrai, A.K., Manrai, L.A. (1994). Latent class multidimensional scaling: A review of recent developments in the marketing and psychometric literature. In: R. P. Bagozzi (Ed.), *Advanced Methods of Marketing Research* (pp. 190-222). Cambridge: Blackwell.
- DeSarbo, W.S., Ramaswamy, V., Chatterjee, R. (1995). Analyzing constant-sum multiple criterion data: A segment-level approach. *Journal of Marketing Research*, 32 (2), 222–232.
- DeSarbo, W. S., Atalay, A.S., Blanchard, S.J. (2009). A three-way clusterwise multidimensional unfolding procedure for the spatial representation of context dependent preferences. *Computational Statistics and Data Analysis*, 53, 3217–3230.
- De Soete, G., Carroll, J.D. (1994). K-Means clustering in a low dimensional Euclidean space. In E. Diday, et al. (Eds.), *New Approaches in Classification and Data Analysis* (pp. 212-219) Heidelberg: Springer.
- Escofier, B., Pagès, J. (1984). Analyse factorielle multiple'. Cahiers du BURO, 2. ISUP. Paris.
- Escofier, B., Pagès, J. (1990). *Analyses Factorielles Simples et Multiples: Objectifs, Méthodes et Interprétation*. Paris: Dunod.
- Fernández-Gómez, M. J., Galindo, M. P., Vicente-Villardón, J. L. (2011). Constrained Unfolding Analysis Based on a Gaussian Response Model for Gradient Analysis in Ecological Studies. Departamento de Estadística, Universidad de Salamanca. Pendiente de publicación.
- Flury, B. (1988). *Common Principal Components and Related Multivariate Models*. New York: Wiley.
- Flury, B. (1995). Developments in principal component analysis. In W. J. Krzanowski (Ed.), *Recent Advances in Descriptive Multivariate Analysis* (pp. 90-112). Oxford: Clarendon Press.
- Gower, J. C. (1975). Generalized Procrustes analysis'. *Psychometrika*, 40, 33-51.
- Green, P. E., & Rao, V. (1972). *Applied Multidimensional Scaling*. Hinsdale, IL: Dryden.
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139–150.
- Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *Annals of Mathematical Statistics*, 17, 144–163.

- Harshman, R. A. (1970). Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 16, 1-84.
- Harshman, R. A., Lundy, M. E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder Jr., J. A. Hattie & R. P. McDonald (Eds.), *Research Methods for Multimode Data Analysis* (pp. 122-215). New York: Praeger.
- Heiser, W. J. (1981). Unfolding analysis of proximity data. Unpublished doctoral dissertation, Leiden University.
- Heiser, W. J. (1985). Multidimensional scaling by optimizing goodness-of-fit to a smooth hypothesis (Tech. Rep. No. RR-85-07). Leiden, The Netherlands: Department of Data Theory, Leiden University.
- Heiser, W. J. (1987). Reihenfolgeninvariante entfaltungsanalyse unter glättebedingungen. *Zeitschrift für Sozialpsychologie*, 18, 220–235.
- Heiser, W. J. (1989). Order invariant unfolding analysis under smoothness restrictions. In G. De Soete, H. Feger, & K. C. Klauer (Eds.), *New Developments in Psychological Choice Modeling* (pp. 3-31). Amsterdam: Elsevier.
- Heiser, W. J. & de Leeuw, J. (1979a). How to use SMACOF-III: A program for metric multidimensional unfolding (Tech. Rep.). Leiden, The Netherlands: Department of Data Theory, Leiden University.
- Heiser, W. J. & de Leeuw, J. (1979b). Metric multidimensional unfolding. *Methoden en Data Nieuwsbrief van de Sociaal Wetenschappelijke Sectie van de Vereniging voor Statistiek*, 4, 26–50.
- Heiser, W. J., Groenen, P. J. F. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minimum. *Psychometrika*, 62 (1), 63–83.
- Kemeny, J. (1959). Mathematics without numbers. *Daedalus*, 88, 571-591.
- Kiers, H. A. L., Vicari, D., Vichi, M. (2005). Simultaneous classification and multidimensional scaling with external information. *Psychometrika*, 70 (3), 433–460.
- Kim, C. (1990). *NEWFOLD: A new unfolding methodology*. Unpublished doctoral dissertation, Department of Marketing, The Wharton School, University of Pennsylvania.
- Kim, C., Rangaswamy, A., & DeSarbo, W. S. (1999). A quasi-metric approach to Multidimensional unfolding for reducing the occurrence of degenerate solutions. *Multivariate Behavioral Research*, 34, 143–180.
- Krzanowski, W. J. (1979). Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74(367), 703-707.
- Krzanowski, W. J. (1982). Between-group comparison of principal components. Some sampling results. *Journal of Statistical Computation and Simulation*, 15, 141-154.
- Kroonenberg, P. M. (1983). *Three-Mode Principal Components Analysis. Theory and Applications*. Leiden, The Netherlands: DSWO-Press.

- Kroonenberg, P. M., de Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, *45*, 69-97.
- Kruskal, J. B., & Carmone, F. J. (1969). *How to use M-D-SCAL (Version 5M) and other useful information* (Tech. Rep.). Murray Hill, N. J.: Bell Telephone Laboratories.
- Kruskal, J. B., & Carroll, J. D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (Vol. 2, pp.639–671). New York: Academic Press.
- Leach, C., Freshwater, K., Aldridge, J., Sunderland, J. (2001). Analysis of repertory grids in clinical practice. *British Journal of Clinical Psychology*, *40*, 225-248.
- Lewin, K. (1951). *Field Theory in Social Science: Selected Theoretical Papers*. New York: Harper.
- L'Hermierdes Plantes, H. (1976). *Structuration des Tableaux a Trois Indices de la Statistique: Theorie et Application d'une Méthode d'Analyse Conjointe*. PhD Thesis, Université des Sciences et Techniques du Languedoc.
- Lingoes, J. C. (1966). An IBM-7090 program for Guttman-Lingoes smallest space Analysis-II. *Behavioral Science*, *11*, 322.
- Lingoes, J. C., & Roskam, E. E. Ch. I. (1973). A mathematical and empirical analysis of two multidimensional scaling algorithms. *Psychometrika*, *38* (Monograph Supplement)
- Mardia, K.V., Kent, J.T. & Bibby, J.M. (1980). *Multivariate Analysis*. New York: Academic Press.
- Martín-Casado, A.M, Galindo, M.P. (1994). Análisis de proximidades: desarrollo, progresos y aplicaciones. *Estadística Española*, *36(135)*, 99-114.
- Martín-Rodríguez, J. (1996). *Contribuciones a la Integración de Subespacios desde una Perspectiva Biplot*. Tesis Doctoral. Departamento de Estadística y Matemática Aplicadas. Universidad de Salamanca.
- Martín-Rodríguez, J., Galindo, M. P., Vicente-Villardón, J. L. (2001). Comparison and integration of subspaces from a biplot perspective. *Journal of Statistical Planning and Inference*, *102(2)*, 1-13.
- Meulman, J.J., Verboon, P. (1993). Points of view analysis revisited: Fitting multidimensional structures to optima distance components with cluster restrictions on the variables. *Psychometrika*, *58 (1)*, 7–35.
- Park, J., Rajagopal, P., DeSarbo, W. (2012). A new heterogeneous multidimensional unfolding procedure. *Psychometrika*, *77 (2)*, 263-287.
- Payne, J.W. (1982). Contingent decision behavior. *Psychological Bulletin*, *92(2)*, 382–402.
- Payne, J.W., Bettman, J.R., Johnson, E.J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43(1)*, 87–131.
- Poole, K.T. (1984). Least Squares Metric Unidimensional Unfolding. *Psychometrika*, *49(3)*, 311-323.

- Puto, C.P. (1987). The framing of buying decisions. *Journal of Consumer Research*, 14(3), 301–315.
- Roskam, E. E. Ch. I. (1968). *Metric Analysis of Ordinal Data*. Voorschoten: VAM.
- Roskam, E. E. Ch. I., & Lingoes, J. C. (1970). MINISSA-I: a FORTRAN IV (G) program for the smallest space analysis of square symmetric matrices. *Behavioral Sciences*, 15, 204–205.
- Schönemann, P. H. (1970). On metric multidimensional unfolding. *Psychometrika*, 35, 349–366.
- Slater, P. (1960). The analysis of personal preferences. *British Journal of Statistical Psychology*, 13, 119–135.
- Steverink, M., Van der Kloot, W. & Heiser, W. J. (2002). *Avoiding degenerate solutions in multidimensional unfolding by using additional distance information* (Tech.Rep. No. PRM 2002-01). Leiden University, the Netherlands: Department of Psychometrics.
- Takane, Y., Young, F.W. & de Leeuw, J. (1977). Nonmetric individual differences MDS: An alternating least squares method with optimal scaling features. *Psychometrika*, 42, 7–67.
- Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications* (pp- 155-167). New York: Wiley.
- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279-311.
- Tucker, L. R. (1972). Relation between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 37, 3-27.
- Tversky, A., Simonson, I. (1993). Context-dependent preferences. *Management Science*, 39(10), 1179–1189.
- Van Buuren, S., Heiser, W.J. (1989). Clustering objects into groups under optimal scaling of variables. *Psychometrika*, 54 (4), 699–706
- Van Deun, K., Heiser, W.J, Debelke, L. (2007). Multidimensional unfolding by nonmetric multidimensional scaling of Spearman distances in the extended permutation polytope. *Multivariate Behavioral Research*, 42(1), 103-132.
- Vera, J.F., Macías, R., Heiser, W.J. (2009). A dual latent class unfolding model for two-way two-mode preference rating data. *Computational Statistics and Data Analysis*, 53, 3231-3244.
- Vichi, M., Kiers, H.A.L. (2001) Factorial k-means analysis for two-way data. *Computational Statistics and Data Analysis*, 37(1), 49–64.
- Wang, M., Schönemann, P. H., Rusk, J. (1975). A conjugate gradient algorithm for the multidimensional analysis of preference data. *Multivariate Behavioral Research*, 10, 45-79.
- Wedel, M., DeSarbo, W.S. (1996). An exponential-family multidimensional scaling mixture methodology. *Journal of Business and Economic Statistics*, 14(4), 447–459.

Young, F. W. & Torgerson, W. S. (1967). TORSCA, a FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science*, 12, 498.

Zarborski, A. (2010). Geometrical presentation of preferences by using unfolding models. *Acta Universitatis Lodzianensis, Folia Oeconomica*, 235, 366-373.